

JOURNAL OF COMPUTATIONAL ELECTRONICS

Proceedings of the
8th INTERNATIONAL WORKSHOP ON
COMPUTATIONAL ELECTRONICS (IWCE-8)
Beckman Institute, University of Illinois
October 15–18, 2001

[20030213 107]

DISTRIBUTION STATEMENT A

Approved for Public Release
Distribution Unlimited

Volume 1—2002

2002 KLUWER ACADEMIC PUBLISHERS
Boston/U.S.A. Dordrecht/Holland London/U.K.

Journal of Computational Electronics

EDITORS

D. K. Ferry
Department of Electrical Engineering
Arizona State University
Box 875706
Tempe, AZ 85287-5706
ferry@asu.edu

K. Hess
Beckman Center
University of Illinois
405 North Mathews Avenue
Urbana, IL 61801
k-hess@uiuc.edu

EDITORIAL BOARD

Narayan Aluru
Beckman Center
University of Illinois

Asen Asenov
Department of Electrical Engineering
University of Glasgow

John R. Barker
Department of Electrical Engineering
University of Glasgow
Glasgow, United Kingdom

Robert Eisenberg
Bard Professor and Chairman
Department of Molecular Biophysics
Rush Medical Center

Stephen Goodnick
Department of Electrical Engineering
Arizona State University

Chihiro Hamaguchi
Department of Electrical Engineering
Kochi Inst. Tech.

Joseph W. Jerome
Department of Mathematics
Northwestern University

Paolo Lugli
Dipartimento di Ingegneria Elettronica
II Universita di Roma tor Vergata

Peter Markowich
Institute for Mathematics
University of Vienna

Wolfgang Porod
Department of Electrical Engineering
Notre Dame University

Umberto Ravaioli
Beckman Center
University of Illinois

Christian Ringhofer
Department of Mathematics
Arizona State University

Gerhard Wachutka
Institute of Physics of
Electrotechnology
Technische Universitäts München

Wolfgang Windl
Department of Material Science
Ohio State University
Columbus, Ohio

JOURNAL OF COMPUTATIONAL ELECTRONICS

Volume 1, Numbers 1/2, July 2002

**Proceedings of the 8th INTERNATIONAL WORKSHOP ON COMPUTATIONAL ELECTRONICS (IWCE-8),
Beckman Institute, University of Illinois, October 15–18, 2001**

Editorial	<i>Umberto Ravaioli</i>	7
Eigenstate Selection in Open Quantum Dot Systems: On the True Nature of Level Broadening	<i>R. Akis, D.K. Ferry and J.P. Bird</i>	9
On the Completeness of Quantum Hydrodynamics: Vortex Formation and the Need for Both Vector and Scalar Quantum Potentials in Device Simulation	<i>John R. Barker</i>	17
On the Current and Density Representation of Many-Body Quantum Transport Theory	<i>John R. Barker</i>	23
A Space Dependent Wigner Equation Including Phonon Interaction	<i>M. Nedjalkov, H. Kosina, R. Kosik and S. Selberherr</i>	27
RTD Relaxation Oscillations, the Time Dependent Wigner Equation and Phase Noise	<i>H.L. Grubin and R.C. Buggeln</i>	33
Modeling of Shallow Quantum Point Contacts Defined on AlGaAs/GaAs Heterostructures: The Effect of Surface States	<i>G. Fiori, G. Iannaccone and M. Macucci</i>	39
Study of Noise Properties in Nanoscale Electronic Devices Using Quantum Trajectories	<i>Xavier Oriols, Ferran Martín and Jordi Suñé</i>	43
Monte-Carlo Simulation of Clocked and Non-Clocked QCA Architectures	<i>L. Bonci, M. Gattobigio, G. Iannaccone and M. Macucci</i>	49
A Wigner Function Based Ensemble Monte Carlo Approach for Accurate Incorporation of Quantum Effects in Device Simulation	<i>L. Shifren and D.K. Ferry</i>	55
The Effective Potential in Device Modeling: The Good, the Bad and the Ugly	<i>D.K. Ferry, S. Ramey, L. Shifren and R. Akis</i>	59
Wigner Paths for Quantum Transport	<i>Paolo Bordone and Carlo Jacoboni</i>	67
Parallelization of the Nanoelectronic Modeling Tool (NEMO 1-D) on a Beowulf Cluster	<i>Gerhard Klimeck</i>	75
Towards Fully Quantum Mechanical 3D Device Simulations	<i>M. Sabathil, S. Hackenbuchner, J.A. Majewski, G. Zandler and P. Vogl</i>	81
Simulation of Field Coupled Computing Architectures Based on Magnetic Dot Arrays	<i>György Csaba and Wolfgang Porod</i>	87

Numerical Acceleration of Three-Dimensional Quantum Transport Method Using a Seven-Diagonal Pre-Conditioner	<i>David Z.-Y. Ting, Ming Gu, Xuebin Chi and Jianwen Cao</i>	93
Numerical Investigation of Shot Noise between the Ballistic and the Diffusive Regime	<i>M. Macucci, G. Iannaccone and B. Pellegrini</i>	99
On Ohmic Boundary Conditions for Density-Gradient Theory	<i>M.G. Ancona, D. Yergeau, Z. Yu and B.A. Biegel</i>	103
Molecular Devices Simulations Based on Density Functional Tight-Binding	<i>Aldo Di Carlo, Marieta Gheorghe, Alessandro Bolognesi, Paolo Lugli, Michael Sternberg, Gotthard Seifert and Thomas Frauenheim</i>	109
Role of Carrier Capture in Microscopic Simulation of Multi-Quantum-Well Semiconductor Laser Diodes	<i>M.S. Hybertsen, B. Witzigmann, M.A. Alam and R.K. Smith</i>	113
Numerical Study of Minority Carrier Induced Diffusion Capacitance in VCSELs Using Minilase	<i>Yang Liu, Fabiano Oyafuso, Wei-Choon Ng and Karl Hess</i>	119
Quantum Transport Simulation of Carrier Capture and Transport within Tunnel Injection Lasers	<i>Wanqiang Chen, Xin Zheng, Leonard F. Register and Michael Stroscio</i>	123
Modeling of Semiconductor Optical Amplifiers	<i>Andrea Reale and Paolo Lugli</i>	129
Hybrid LSDA/Diffusion Quantum Monte-Carlo Method for Spin Sequences in Vertical Quantum Dots	<i>P. Matagne, T. Wilkens, J.P. Leburton and R. Martin</i>	135
Theoretical Investigations of Spin Splittings and Optimization of the Rashba Coefficient in Asymmetric AlSb/InAs/GaSb Heterostructures	<i>X. Cartoixà, D.Z.-Y. Ting and T.C. McGill</i>	141
Modeling Spin-Dependent Transport in InAs/GaSb/AlSb Resonant Tunneling Structures	<i>D.Z.-Y. Ting, X. Cartoixà, T.C. McGill, D.L. Smith and J.N. Schulman</i>	147
Tunneling through Thin Oxides—New Insights from Microscopic Calculations	<i>M. Städele, B. Tuttle, B. Fischer and K. Hess</i>	153
Full Quantum Simulation of Silicon-on-Insulator Single-Electron Devices	<i>Frederik Ole Heinz, Andreas Schenk, Andreas Scholze and Wolfgang Fichtner</i>	161
A 3-D Atomistic Study of Archetypal Double Gate MOSFET Structures	<i>Andrew R. Brown, Jeremy R. Watling and Asen Asenov</i>	165
3-D Parallel Monte Carlo Simulation of Sub-0.1 Micron MOSFETs on a Cluster Based Supercomputer	<i>Asim Kepkep and Umberto Ravaioli</i>	171
Hole Transport in Orthorhombically Strained Silicon	<i>F.M. Bufler and W. Fichtner</i>	175
Empirical Pseudopotential Method for the Band Structure Calculation of Strained-Silicon Germanium Materials	<i>Salvador Gonzalez, Dragica Vasileska and Alexander A. Demkov</i>	179
A Computational Exploration of Lateral Channel Engineering to Enhance MOSFET Performance	<i>Jing Guo, Zhibin Ren and Mark Lundstrom</i>	185
Monte Carlo Simulations of Hole Dynamics in Si/SiGe Quantum Cascade Structures	<i>Z. Ikoníć, P. Harrison and R.W. Kelsall</i>	191
Calculation of Direct Tunneling Current through Ultra-Thin Gate Oxides Using Complex Band Models For SiO ₂	<i>Atsushi Sakai, Akihiro Ishida, Shigeyasu Uno, Yoshinari Kamakura, Masato Morifuji and Kenji Taniguchi</i>	195

Comparison of Quantum Corrections for Monte Carlo Simulation	Brian Winstead, Hideaki Tsuchiya and Umberto Ravaioli	201
Monte Carlo Based Calculation of the Electron Dynamics in a Two-Dimensional GaN/AlGaN Heterostructure in the Presence of Strain Polarization Fields	Tsung-Hsing Yu and Kevin F. Brennan	209
Parallel Approaches for Particle-Based Simulation of Charge Transport in Semiconductors	M. Saraniti, J. Tang, S. Goodnick and S. Wigger	215
Full-Band Monte Carlo Simulation of Two-Dimensional Electron Gas in SOI MOSFETs	H. Takeda, N. Mori and C. Hamaguchi	219
Band-to-Band Tunneling by Monte Carlo Simulation for Prediction of MOSFET Gate-Induced Drain Leakage Current	Edwin C. Kan, Venkat Narayanan and Gen Pei	223
A Computational Technique for Electron Energy States Calculation in Nano-Scopic Three-Dimensional InAs/GaAs Semiconductor Quantum Rings Simulation ... Yiming Li, O. Voskoboynikov, C.P. Lee and S.M. Sze		227
Fully Numerical Monte Carlo Simulator for Noncubic Symmetry Semiconductors	Louis Tirino, Michael Weber, Kevin F. Brennan, Enrico Bellotti, Michele Goano and P. Paul Ruden	231
Theoretical Study of RF Breakdown in GaN Wurtzite and Zincblende Phase MESFETs	M. Weber, L. Tirino, K.F. Brennan and Maziar Farahmand	235
Quantum Mechanical Model of Electronic Stopping Power for Ions in a Free Electron Gas Yang Chen, Di Li, Geng Wang, Li Lin, Stimit Oak, Gaurav Shrivastav, Al F. Tasch and Sanjay K. Banerjee	241
An Analytical 1-D Model for Ion Implantation of Any Species into Single-Crystal Silicon Based on Legendre Polynomials	G. Shrivastav, D. Li, Y. Chen, G. Wang, L. Lin, S. Oak, A.F. Tasch and S.K. Banerjee	247
On the Electron Transient Response in a 50 nm MOSFET by Ensemble Monte Carlo Simulation in Presence of the Smoothed Potential Algorithm	Gabriele Formicone, Marco Saraniti and David K. Ferry	251
Quantum Corrections in the Monte Carlo Simulations of Scaled PHEMTs with Multiple Delta Doping	K. Kalna and A. Asenov	257
Thermally Self-Consistent Monte Carlo Device Simulations	N.J. Pilgrim, W. Batty and R.W. Kelsall	263
3D Monte Carlo Modeling of Thin SOI MOSFETs Including the Effective Potential and Random Dopant Distribution	S.M. Ramey and D.K. Ferry	267
Low-Field Mobility and Quantum Effects in Asymmetric Silicon-Based Field-Effect Devices	I. Knezevic, D. Vasileska, X. He, D.K. Schroder and D.K. Ferry	273
Quantum Potential Corrections for Spatially Dependent Effective Masses with Application to Charge Confinement at Heterostructure Interfaces	J.R. Watling, J.R. Barker and S. Roy	279
Comparison of Three Quantum Correction Models for the Charge Density in MOS Inversion Layers	Xinlin Wang and Ting-wei Tang	283
Can the Density Gradient Approach Describe the Source-Drain Tunnelling in Decanano Double-Gate MOSFETs?	J.R. Watling, A.R. Brown and A. Asenov	289
A Particle Description Model for Quantum Tunneling Effects	Hideaki Tsuchiya and Umberto Ravaioli	295

Journal of Computational Electronics is published quarterly.

SUBSCRIPTION RATES

The subscription price of *Journal of Computational Electronics* for 2002, Volume 1 (4 quarterly issues), including postage and handling is:

Print OR Electronic Version: EURO 350.00/US \$350.00 per year

ORDERING INFORMATION/ SAMPLE COPIES

Subscription orders and requests for sample copies should be sent to:

Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, MA 02061 USA
phone: (781) 871-6600
fax: (781) 871-6528
e-mail: kluwer@wkap.com

or Kluwer Academic Publishers
P.O. Box 322
3300 AH Dordrecht
The Netherlands

or to any subscription agent

© 2002 Kluwer Academic Publishers

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

Photocopying. *In the U.S.A.:* This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, U.S.A.

Authorization to photocopy items for internal or personal use, or the internal or personal use or specific clients, is granted by Kluwer Academic Publishers for users registered with the Copyright Clearance Center (CCC). The "services" for users can be found on the internet at: www.copyright.com. For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising of promotional purposes, for creating new collective works, or for resale.

In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to Kluwer Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Periodicals postage paid at Rahway, NJ

U.S. mailing agent: Mercury Airfreight International, Ltd.
365 Blair Road
Avenel, NJ 07001 U.S.A.

Published by Kluwer Academic Publishers

Postmaster: Please send all address corrections to: *Journal of Computational Electronics*, c/o Mercury Airfreight International, Ltd., 365 Blair Road, Avenel, NJ 07001, U.S.A.

ISSN: 1569-8025

Printed on acid-free paper



Editorial

It is a great privilege for me to serve as guest editor of the first issue of the Journal of Computational Electronics, and it seemed only natural to make the kick-off of this new journal coincide with the publication of the Proceedings of the 8th International Workshop on Computational Electronics (IWCE), held at the Beckman Institute of the University of Illinois on October 15–18, 2001. Over the last decade, the IWCE has grown into the main forum where new results and ideas in computational electronics are presented and discussed. A national Workshop on Computational Electronics was first held in 1990 at the Beckman Institute, under the auspices of the National Center for Computational Electronics and the National Science Foundation. The goal of the meeting was to foster interdisciplinary interaction between scientists in electrical engineering, physics, applied mathematics and computer science. The experiment proved to be very successful and in 1992 the first IWCE was held, again at the Beckman Institute, followed by events in Leeds, U.K. (1993); Portland, OR (1994); Tempe, AZ (1995); Note Dame, IN (1996); Osaka, Japan (1998); Glasgow, U.K. (2000), and again Urbana, IL (2001). The next workshop will be for the first time in Italy, in 2003. The format of IWCE creates many opportunities for interaction and discussion among the participants, always with a large representation of graduate students who are particularly encouraged to attend and present papers or posters. Many lasting collaborative interactions have resulted from discussions initiated at an IWCE.

The Journal of Computation Electronics fills the need for a publication dedicated to physical simulation of devices and processes, with a focus on interdisciplinary work and large scale supercomputing applications. The community typically attending IWCE best represents the audience addressed by the Journal of Computational Electronics, but the composition of this community has grown over the years to include even more discipline areas. The emphasis of the first workshops was on classical device simulation approaches (drift-diffusion and hydrodynamics models) and particle Monte Carlo methods, while other areas have gained increasing importance at following meetings. These areas include quantum transport and quantum device simulation, opto-electronics, process simulation and, more recently, molecular devices, MEMS and transport in biological ion channels. Rapid technological advances in new directions of research and the widespread availability of high performance computers and clusters, have clearly challenged the computational electronics community to address simulation problems of increasing complexity in the nano-technology area. These efforts require even more contributions from other fields of expertise, from heat transfer and micro-fluidics to computational chemistry and computational biology.

I believe that the Journal of Computational Electronics has the potential to become the pre-eminent publication on multidisciplinary aspects of electronics simulation, with the editorship in the capable hands of David Ferry and Karl Hess, some of the most respected scientists in computational electronics (and incidentally the two people who have been most influential on my own professional career). The membership of the editorial board includes international leaders, covering an impressive range of expertise in all relevant areas. While working on the preparation of the IWCE-8 proceedings issue, I was also very impressed by the high quality of the contributions and I am confident that the Journal of Computational Electronics is off to a good start. If the quality of future submission to the regular issues will continue to be on this level, the success of this new journal is assured.

The quality of the papers submitted for publication on the IWCE proceedings also reflects the commitment by funding agencies and institutions that have continued to support the workshop over the years. IWCE-8 would not have been possible without the direct support of the National Science Foundation, the Beckman Institute of the University of Illinois, the US Office of Naval Research, the Distributed Center for Advanced Electronics Simulation (DesCArES), and the technical sponsorship of the IEEE.

Umberto Ravaioli

University of Illinois at Urbana-Champaign



Eigenstate Selection in Open Quantum Dot Systems: On the True Nature of Level Broadening*

R. AKIS,[†] D.K. FERRY AND J.P. BIRD

*Center for Solid State Electronics Research and Department of Electrical Engineering,
Arizona State University, Tempe, AZ 85287-5706, USA*

richard.akis@asu.edu

Abstract. We show that transport in open quantum dots can be mediated by single eigenstates, even when the leads allow several propagating modes. The broadening of these states, generally localized in the interior, can be virtually independent of lead width. As such, the Thouless argument, invoked to suggest that all states should be unresolvable under these conditions, can in fact fail. Thus, any transport theory based on such assumptions (in particular, random matrix theory) must be called into question, as the fluctuations produced by these states can in fact dominate the conductance. These trapped states also produce interesting and potentially useful effects in coupled dot systems as well.

Keywords: device modeling, quantum transport, quantum dots

1. Introduction

First applied to disordered conductors, the Thouless argument relates the conductance of a system to the diffusion-induced broadening of its energy levels. Accordingly, a metal may be viewed as a system with strongly-broadened energy levels, while an insulator is one whose density of states (DOS) consists of isolated peaks (Thouless 1977). While the Thouless argument provides an understanding of the origins of localization in disordered conductors, it has recently become possible to study electron transport in ballistic quantum dots (Jalabert, Baranger and Stone 1990, Baranger, Jalabert and Stone 1993a, b, Lin and Jensen 1996, Wirtz, Tang and Burgdörfer 1997, Marcus *et al.* 1992, Chang *et al.* 1994, Bird *et al.* 1996, 1999, Sachrajda *et al.* 1998). These open structures consist of a central scattering cavity that is coupled to external reservoirs by means of quantum point contacts (QPCs). Since the conductance of these structures (measured in units of the dimensionless conductance e^2/h) is typically larger than unity, it is often thought that the Thouless argument may be

used to imply that their discrete DOS is unimportant for an understanding of transport. A key feature of the Thouless argument is an assumption of *uniform* level broadening, independent of the specific details of the energy states. While this seems reasonable for diffusive conductors, in open quantum dots we demonstrate here that the level broadening is highly non-uniform and that single eigenstates may remain resolved, thus demonstrating that the Thouless argument does not generally hold. These results have important implications for theoretical analyses of such structures.

This paper is organized as follows. In Section 2, the Thouless argument is summarized. In Section 3, our method of calculation is briefly described. A discussion on conductance resonances in open dots is found in Section 4. In Section 5, we discuss decomposing the open dot wave functions in terms of closed dot eigenstates. In Section 6, the focus is shifted to coupled dot systems. Conclusions are drawn in Section 7.

2. The Thouless Argument

The Thouless argument follows by noting that the energy levels in a conductor of length L should be

*Work supported by the Office of Naval Research.

[†]To whom correspondence should be addressed.

uncertain by an amount $\Gamma \approx hD/L^2$, where D is the diffusion constant and L^2/D is the time required to diffuse across the sample. Since the average level spacing in the conductor may be written as $\Delta = 1/N_E L^d$, where N_E is the DOS and d is the dimensionality, the ratio Γ/Δ may be written as (Lee, Stone and Fukuyama 1987):

$$\frac{\Gamma}{\Delta} = \frac{hD}{L^2} N_E L^d. \quad (1)$$

Using the Einstein relation ($\sigma = e^2 N_E D$) to relate the DOS to the conductivity (σ), Eq. (1) may be simplified to yield:

$$\frac{\Gamma}{\Delta} = \frac{h}{e^2} \sigma L^{d-2} = g, \quad (2)$$

where g is the dimensionless conductance, with units of e^2/h . Equation (2) is the crux of the Thouless argument and suggests that, in a metallic conductor ($g \geq 1$), the level broadening, Γ , is always comparable to, or greater than, the average level spacing. In a dot whose point contacts each support N propagating modes then, by assuming Ohmic addition of the two point contacts, the conductance may be written as $g = N$. Since $N \geq 1$ is required for the dot to be open, it is therefore often argued that Eq. (2) proves that the energy levels of open dots can never be resolved ($\Gamma \geq \Delta$ for $N \geq 1$).

3. Method of Calculation

Our simulations are performed on a discrete lattice using a numerically stabilized variant of the transfer matrix approach (Usuki *et al.* 1995). The dot is enclosed inside a waveguide which extends a finite number of lattice sites in the transverse (y) direction. The structure is broken down into a series of slices along the longitudinal (x) direction. Imposing an electron flux from the left, one translates across successive slices and, on reaching the end, one obtains the transmission coefficients which enter the Landauer-Büttiker formula to give the conductance. In cases where we examine closed dots, to obtain the spectrum and the eigenstates, we solve a finite difference Schrödinger equation with Dirichlet boundary conditions. This sparse matrix eigenvalue problem is done numerically by using ARPACK routines (www.caam.rice.edu/software/ARPACK/index.html), which use Lanczos/Arnoldi factorization.

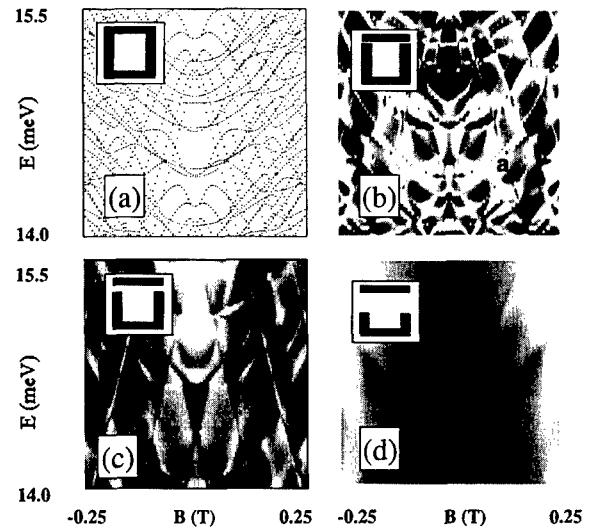


Figure 1. In (a), a portion of the spectrum is plotted as a function of E and B for a $0.3 \mu\text{m}$ square dot. The conductance, G , is plotted vs. energy, E , and magnetic field, B , for open quantum dots with leads allowing (b) one mode, (c) four modes and (d) nine modes. The lighter regions of shading correspond to higher values of the conductance G . The dot schematics are shown in the insets. The labels a and b correspond to the positions of Fano resonances.

4. Conductance Resonances in Open Dots

We begin by showing the correspondence between the energy spectrum of a *closed* square dot with the conductance features exhibited by the open system. Figure 1(a) shows a portion of the energy spectrum as a function of magnetic field for a $0.3 \mu\text{m}$ quantum dot. Figure 1(b) shows what happens when the dot is now opened and connected to external waveguides by QPCs that are at the top edge of the dot as shown in the inset. In this case the width of the QPCs have been adjusted so that a single mode propagates. What is plotted is $G(E, B)$ with lighter shading corresponding to higher conductance. For the entire energy range shown in this picture, $G < 2e^2/h$. The picture shows *resonant* behavior, as indicated by the striations that are superimposed on the conductance. Comparing this picture with the spectrum shown in Fig. 1(a), $G(E, B)$ clearly shows the influence of the closed dot DOS, as the basic pattern is reproduced. However, certain resonance lines appear to be shifted in comparison to their spectra counterparts and there are certain features in the conductance that *apparently* do not have a spectral analog. In particular, there are linear resonance features that actually cross at $B = 0 \text{ T}$. In contrast, the spectrum shows lines that appear linear for much of the range

shown, but bend over in the region near $B = 0$ T. Thus, rather than crossing, they appear to form a type of *anti-crossing*. This line shifting and line creation illustrates another effect that the QPCs have—they act as a perturbation that results in the creation of *new* eigenstates not present in the perfectly square system. In Fig. 1(c), the QPCs have now been adjusted to permit four modes to propagate. The conductance here ranges from $\sim 2e^2/h$ to $\sim 8e^2/h$. Despite the fact that the dot is far more “open” than in the previous case there is still resonant behavior. However, the picture is somewhat simplified compared to Fig. 1(b). What remains are a series of parabolic curves as well as sets of almost parallel resonance lines, tilted to the left and the right, forming a very regular cross hatched pattern. These patterns yield characteristic fingerprints in the conductance fluctuations that have in fact been observed experimentally (Bird, Akis and Ferry 1999). In Fig. 1(d), the QPCs support nine modes. Here the parallel lines have vanished, leaving only the parabolic striations. Clearly the broadening introduced by the QPCs is highly nonuniform.

In Fig. 2(a) and (b), respectively, we plot $G(E)$ vs. E , focusing on the conductance resonances labeled “a” and “b” in the previous figure. The asymmetric line-shape of these features is characteristic of Fano resonances, which occur in systems where quasi-bound states are coupled to a continuum (see Göres *et al.* (2000) and references therein). These may be represented by the functional form (Göres *et al.* 2000):

$$G = G_b + \frac{G_0(\varepsilon + q)^2}{\varepsilon^2 + 1}, \quad (3)$$

where $\varepsilon = (E - E_R)/\Gamma$, E_R is the energy on resonance, q is an asymmetry parameter that depends on the background phase shift, G_b the background conductance that the resonance sits upon, and G_0 determines the magnitude. The dashed lines are the fits. *Significantly, the resonance in (b) is sharper than in (a), even though the QPCs are much wider.* The insets show the resonant wave functions which are both of the “bouncing ball” variety. That is, the standing waves trapped between the upper and lower boundaries appear to be aligned with the orbital trajectory that a classical billiard would take if it were bouncing between them. This behavior is reminiscent of the *scarring* of the wave function by classical orbits observed in chaos theory (Heller 1984). The two resonant states shown here can be thought of as being largely equivalent. The resonance in (b) however occurs at a lower

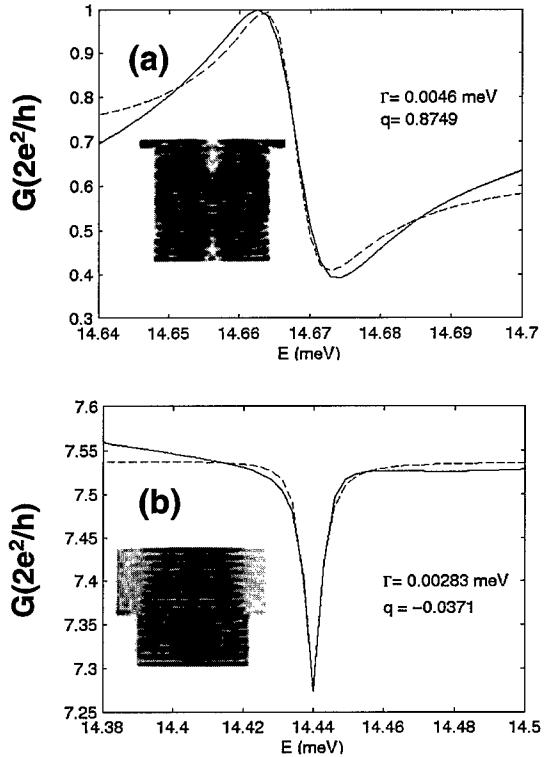


Figure 2. Conductance, G vs. energy, E , for the single mode dot (a) and the nine mode dot (b). Fits to the Fano resonance formula are also plotted as dotted lines.

energy because the “effective” dot size is larger. This point is explored in further detail in the next section.

5. Eigenstate Decomposition

The relationship between the open dot resonances and the eigenstates of the corresponding closed system can be quantified by doing a spectral decomposition. Since closed-dot eigenstates form an orthogonal basis set, the wave functions of the open dot can be expressed as a linear combination of these states by means of projection in the dot region:

$$\psi = \sum_n C_n \phi_n^{closed}, \quad \text{where } C_n = \langle \psi | \phi_n^{closed} \rangle. \quad (4)$$

Figure 3(a) shows $G(E)$ vs. E for a nearly square dot (the dimensions are $0.3 \mu\text{m}$ by $b = 0.307 \mu\text{m}$, the non-commensurate shape was chosen to insure that the levels of the rectangular dot were not degenerate). The QPCs allow 2 modes in the energy range displayed. At the top are markers that indicate the positions of the

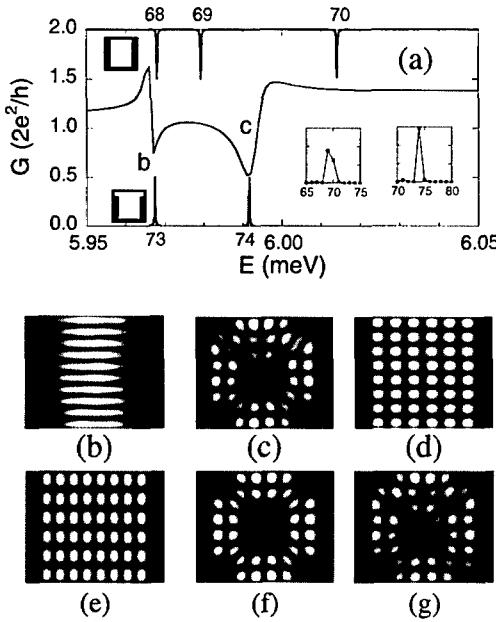


Figure 3. (a) $G(E)$ vs. E for the rectangular dot. The marks at the top indicate the positions of the eigenenergies for the closed rectangle, and the marks at the bottom are for the T-shaped cavity. Left inset: the rectangular decomposition. Right inset: the T decomposition. (b) $|\Psi(x, y)|$ vs. x and y , the wave function in the interior region of the open dot $E = 5.988$ meV. (c) As in (b) but for $E = 5.9915$ meV. (d) The 69th eigenstate of the rectangle. (e) The 70th eigenstate of the rectangle. (f) A linear combination of the states shown in (d) and (e). (g) The 74th eigenstate of the T-shaped cavity.

68th through 70th eigenenergies for closed rectangular dot. At the bottom are markers for the 73rd and 74rd eigenenergies for a second type of dot in the form of a T-shaped cavity (note the inset in the bottom left corner). The conductance over this energy range shows two major resonances, the first of which, at $E = 5.988$ meV (marked b) lines up with both rectangular and T eigenstates. Figure 3(b) shows the corresponding open dot wave function, which happens to closely resemble the 68th rectangular state and the 73rd T state. The second resonance, at $E = 5.9915$ meV, marked c, lines up only with a T state. Figure 3(c) shows the corresponding open dot wave function. The left inset shows the decomposition of this wave function in terms of rectangular states. Two states, $n = 69$ (Fig. 3(d)) and $n = 70$ (Fig. 3(e)), which bracket the resonance, yield the vast majority of the total. The linear combination these two states produces (f), which is virtually identical to (c) in the interior region. If we stopped here, one might conclude that, despite the presence of a resonance, the fact that a number of eigenstates contribute indicates

that the level structure of the dot is not preserved. This however is wrong, because the rectangular cavity is not the appropriate system for comparison. This is illustrated by the right inset of (a), which gives the T state decomposition. Here, a single state, $n = 74$ (panel (g)) of the T cavity yields the vast majority of the amplitude. Comparing (g) with (c), it is difficult to pick out the open state from the closed one. The 74th T state can be viewed as a hybridized state resulting from a perturbation (the extensions added onto the sides to mimic the QPCs) which has mixed the 69th and 70th rectangular states together. A very important property of this state is that, despite the fact that it results from a T geometry, the amplitude is almost entirely concentrated away from the perturbing leads. The 74th state survives in the open system precisely because of its locality and apparent disconnection from the QPCs, an ironic result since the QPCs provided the perturbation that created it.

Quantum dots have generated much interest as a test bed for the study of quantum chaos (Jalabert, Baranger and Stone 1990; Baranger, Jalabert and Stone 1993a, b, Lin and Jensen 1996, Wirtz, Tang and Burgdörfer 1997, Marcus *et al.* 1992, Chang *et al.* 1994). It has been predicted that certain physical properties should depend on whether the dot has a geometry with classically regular behavior (e.g. the rectangle) or a geometry that induces classical chaos, such as the stadium. However, with regards to this resonance phenomena, the stadium actually behaves in a very similar manner to a rectangular dot.

This is illustrated by Fig. 4(a), which shows the conductance for an open stadium quantum dot. Here the system is very open- the width of the QPCs is 60% of the breadth of the dot. The energy here is normalized to the average level spacing ($\Delta = 2\pi\hbar^2/m^*A$, where A is the stadium area). The squares represent the energy levels of the standard stadium, and the triangles those of a perturbed stadium, as shown in the lower right inset. In both cases, twelve eigenvalues lie in the plotted energy range. In contrast, G exhibits only three well-defined resonances over this same range, which we label with the indices (i)–(iii). All three resonances line up in energy with eigenstates of the perturbed stadium, while resonances (ii) shows no correspondence to a standard stadium eigenvalue. The perturbed stadium states in question, the 134th, 138th and 142nd, are shown as insets in Fig. 5(b). As with the previous example, these surviving states are all scarred by “bouncing ball” orbits with amplitude concentrated away from the QPCs. As one might expect, the states that have amplitude

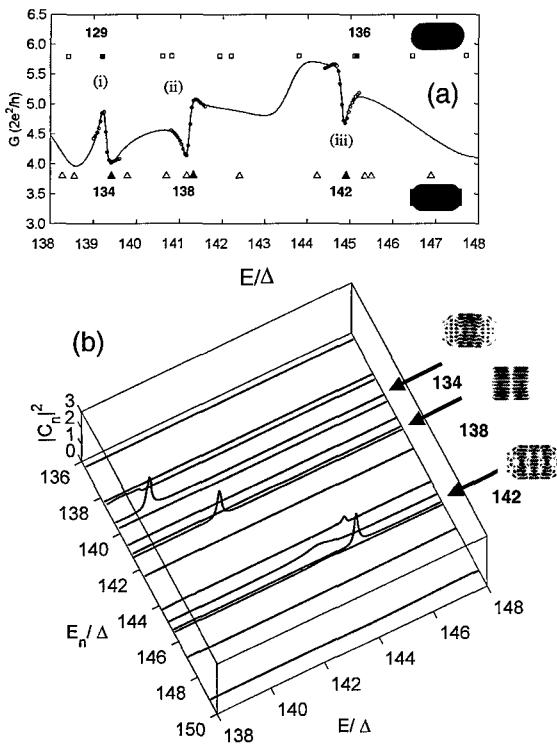


Figure 4. (a) $G(E)$ vs. E/Δ for an open stadium with six modes in the QPCs. The squares and triangles represent the energies of the eigenstates of the unperturbed and perturbed stadium, respectively. The circles are fits to the Fano formula. (b) The perturbed-stadium decomposition coefficients $|C_n|^2$ vs. E/Δ and E_n/Δ . The 134th, 138th and 142nd eigenstates are shown as insets.

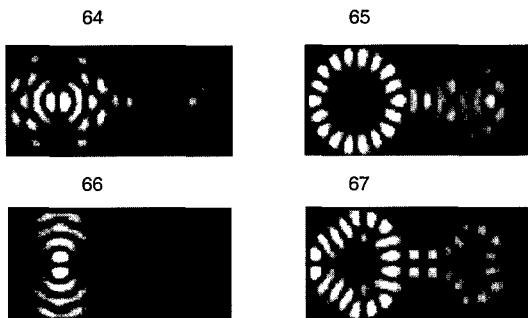


Figure 5. States 64 through 67 of an asymmetric coupled dot system.

near the outer perimeter of the dot do not survive when the dot is opened up.

By fitting to the Fano formula, one can obtain the level widths, which we find to be only a small fraction of the average level spacing in each case ($\Gamma = 0.075\Delta$ (i), 0.097Δ (ii), and 0.104Δ (iii)). It should be noted that the values one obtains for the level widths of these

particular states is remarkably stable over a large range of QPC openings.

In Fig. 4(b), we plot the decomposition $|C_n|^2$ vs. E/Δ and E_n/Δ . We have included the E_n/Δ axis and made this a three-dimensional plot to call attention to the actual spacing of the energy levels. Significantly, we see that the width of the 134th state, as inferred from the breadth of the decomposition peak along the E/Δ axis, is larger than the spacing between the 134th and 135th levels. Similar observations may also be made concerning the other resonant states. Nonetheless, the decompositions remain dominated by the contribution of single eigenstates. Contrary to naive assumptions, the mere proximity of levels in energy need not be an obstacle to being able to resolve individual states in open dots. In a one-dimensional quantum-well problem, lowering the barriers leads to broadening of states in a simple and predictable manner. In two-dimensional quantum dots, the situation is far more complex. The geometry of the system plays as important role in determining the level broadening as the size of the QPC opening itself.

6. Coupled Dot Systems

QPCs, as we have shown, can generate resonant states with amplitude localized or concentrated in particular regions of a dot, these states resulting from a mixing of unperturbed dot states. When two or more dots are coupled together, one expects a similar process to take place, whereby “atomic”, single-dot, states become hybridized to generate the “molecular” states of the coupled system.

In Fig. 5, we show states 64–67 of an asymmetric dot system, with the right dot having a smaller radius. These results clearly show that the coupled system can show a combination of behavior-coupled dot states that have strong single dot characteristics (e.g. state 66, and, to a lesser extent, state 65), as well as states where the two dots truly act collectively as one unit (e.g. state 64). These results suggest that the transition from “atomic” to “molecular” behavior is not a simple one and there can be an intermingling of these regimes. In the past, it has been suggested that, once the QPC is wide enough to support a single mode, the coupled system essentially behaves as if it were simply one large single dot (Livermore *et al.* 1996). Here, the connecting QPC supports 2 modes, well beyond the tunneling regime. The fact that the 66th state shown here has almost all its amplitude concentrated in one dot suggests interesting

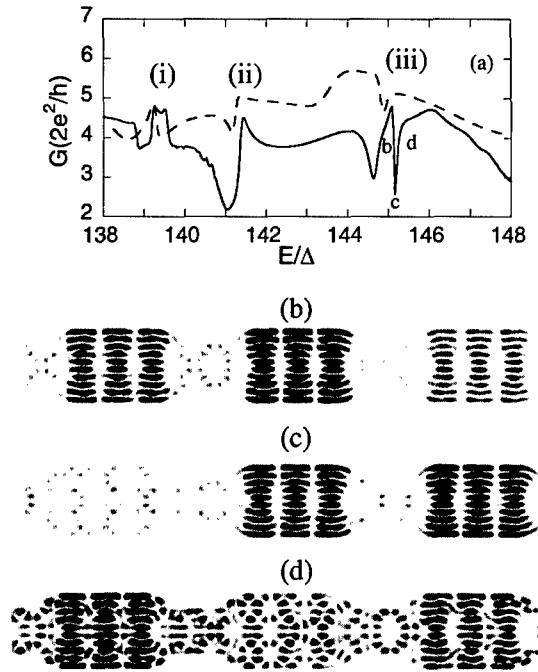


Figure 6. (a) $G(E)$ vs. E/Δ for a three stadium chain is plotted (solid line). The dashed line is the single stadium result from Fig. 5(a). The labels a, b and c correspond the energies of the wave functions shown in panels (b), (c) and (d) respectively.

practical possibilities. One can engineer a coupled system whereby states close in energy each have this amplitude localization, but in different dots.

In Fig. 6, we illustrate this using an open chain of three coupled stadium dots. Each of the individual stadiums is identical to that used in the previous section. Comparing the single dot conductance with that of the chain, one sees that resonances (i) and (iii) have been split into multiple resonances, while (ii) has become deeper and wider. If more dots were added, these features would ultimately correspond to the formation of bands and gaps (Leng and Lent 1993, 1994). The wave functions shown in panels (b), (c) and (d) show the switching behavior alluded to above, with the amplitude of the wave function being switched between individual dots depending on the energy. In an actual experimental realization of such a system, the system parameters that would be tuned to achieve such effects could be the gate voltage or an applied magnetic field.

7. Conclusions

We have demonstrated that level quantization is preserved in the open dots, but is done so selectively. With

regards to the selection of particular states, the QPCs perform this task by “anti-selection”—certain closed dot states are not allowed in the open system. Scarred states in particular tend to survive because their amplitude is localized in certain dot regions. The QPCs also act as a perturbation which creates new states by hybridization. A model that assumes uniform level broadening cannot provide an accurate general description for the physics of open dots and so the Thouless argument cannot really be applied. This result has important implications. In particular, the RMT based semi-classical approach (Jalabert, Baranger and Stone 1990, Baranger, Jalabert and Stone 1993a, b, Lin and Jensen 1996, Wirtz, Tang and Burgdörfer 1997, Marcus *et al.* 1992, Chang *et al.* 1994) commonly applied to open dots has a far more limited range of validity than previously thought, as it assumes a completely broadened spectrum *a priori* and ignores the resonant structure that can actually dominate the conductance fluctuations. It should also be mentioned that scarred resonant states analogous to ours, with amplitude localized in the interior, have also been found in simulations of Coulomb blockaded dots (Silvestrov and Imry 2000). These earlier results combined with those shown here, indicate that there is no simple transition between “closed” and “open” regimes. Any distinction made simply on the basis of mode number and/or average level spacing is a purely arbitrary one.

In closing, it should be noted that most of the effects discussed here (for example, the robust nature of the level widths) are a manifestation of resonance trapping, a phenomenon previously noted in the context of nuclear physics (Muraviev *et al.* 1999) and microwave cavities (Persson *et al.* 2000).

References

- Baranger H.U., Jalabert R.A., and Stone A.D. 1993a. Phys. Rev. Lett. 70: 3876.
- Baranger H.U., Jalabert R.A., and Stone A.D. 1993b. Chaos 3: 665.
- Bird J.P., Akis R., and Ferry D.K. 1999. Phys. Rev. B 60: 13676.
- Bird J.P., Akis R., Ferry D.K., Vasileska D., Cooper J., Aoyagi Y., and Sugano T. 1999. Phys. Rev. Lett. 82: 4691.
- Bird J.P., Ferry D.K., Akis R., Ishibashi K., Aoyagi Y., Sugano T., and Ochiai Y. 1996. Europhys. Lett. 35: 529.
- Chang A.M., Baranger H.U., Pfeiffer L.N., and West K.W. 1994. Phys. Rev. Lett. 73: 2111.
- Göres J., Goldhaber-Gordon D., Heemeyer S., Kastner M.A., Shtrikman H., Mahalu D., and Meirav U. 2000. Phys. Rev. B 62: 2188.
- Heller E.J. 1984. Phys. Rev. Lett. 53: 1515.
- Jalabert R.A., Baranger H.U., and Stone A.D. 1990. Phys. Rev. Lett. 65: 2442.

- Lee P.A., Stone A.D., and Fukuyama H. 1987. Phys. Rev. B 35: 1039.
- Leng M. and Lent C.S. 1993. Phys. Rev. Lett. 71: 137.
- Leng M. and Lent C.S. 1994. Phys. Rev. B 50: 10823.
- Lin W.A. and Jensen R.V. 1996. Phys. Rev. B 53: 3638.
- Livermore C., Crouch C.H., Westervelt R.M., Campman K.L., and Gossard A.C. 1996. Science 274: 5291.
- Marcus C.M., Rimberg A.J., Westervelt R.M., Hopkins P.F., and Gossard A.C. 1992. Phys. Rev. Lett. 69: 506.
- Muraviev S.E., Rotter I., Shlomo S., and Urin M.H. 1999. Phys. Rev. C 59: 2040.
- Persson E., Rotter I., Stöckmann H.-J., and Barth M. 2000. Phys. Rev. Lett. 85: 2478.
- Sachrajda A.S., Ketzmerick R., Gould C., Feng Y., Kelly P.J., Delage A., and Wasilewski Z. 1998. Phys. Rev. Lett. 80: 1948.
- Silvestrov P.G. and Imry Y. 2000. Phys. Rev. Lett. 85: 2565.
- Thouless D.J. 1977. Phys. Rev. Lett. 39: 1167.
- Usuki T., Saito M., Takatsu M., Kiehl R.A., and Yokoyama N. 1995. Phys. Rev. B 52: 8244.
- Wirtz L., Tang J.-Z., and Burgdörfer J. 1997. Phys. Rev. B 55: 7589.



On the Completeness of Quantum Hydrodynamics: Vortex Formation and the Need for Both Vector and Scalar Quantum Potentials in Device Simulation

JOHN R. BARKER

Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8LT, Scotland, UK
jbarker@elec.gla.ac.uk

Abstract. The conditions for the occurrence of quantized vortices in electron flow are examined critically in the context of quantum hydrodynamic modelling. The presence of vortices is shown to be described by the coupling to a new vector quantum potential which augments the conventional scalar quantum potential used in hydrodynamic and density gradient modelling of semiconductor devices.

Keywords: quantum hydrodynamics, vortices, semiconductor theory

1. Introduction

The present interest in decanano semiconductor FETs, open quantum dot structures and prospective quantum computing devices has led to a substantial increase in the use of quantum hydrodynamic, quantum Monte Carlo and Wigner function simulations. These methods have been successfully deployed in one-dimensional problems such as resonant tunnelling devices, although some questions linger over the validity of the quantum potential models (Barker and Ferry 1998), especially the high temperature approximations. Here we describe results from a detailed analytical and numerical study of 2D and 3D quantum transport in semiconductors under conditions in which quantized vortices (Barker, Ferry and Akis 2000, Barker 2001, Lent 1990) may occur.

2. Vortex Formation: Pure State Description

It is well-known that the equations of pure-state quantum hydrodynamics may be derived by taking the polar form for the wavefunction and separating out the real and imaginary parts of the Schrödinger equation. This results in the continuity equation for the amplitude squared $n = R^2$ and a Hamilton Jacobi-like equation (for S the phase of the wavefunction) the

gradient of which leads to an Euler-like equation for the velocity field defined by $\mathbf{v} = \mathbf{J}(\mathbf{r}, t)/n(\mathbf{r}, t)$ (here the particle current density is $J = n(\mathbf{r}, t)\nabla S(\mathbf{r}, t)/m$). There are two differences with classical hydrodynamic models: first, there appears a scalar quantum potential $V_Q = (-\hbar^2/2m)n^{-1/2}\nabla^2\sqrt{n}$ within the Euler equation; secondly, the resulting equations of motion are not complete, there remains an additional constraint imposed by the single-valuedness of the wavefunction leading to the quantization of velocity circulation (Barker and Ferry 1998):

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = Nh/m \quad (1)$$

Any spatial circuit C through which a vortex occurs in the velocity flow leads to a non-zero integer in condition (1). Vortex cores occur along the strong nodal lines of the amplitude $R(\mathbf{r}, t)$ of the wavefunction, where the phase S/\hbar is indeterminate. Here we define a *strong nodal line* $\mathbf{x}_i(\mathbf{r}, t)$ such that in its vicinity $R \approx |\mathbf{r} - \mathbf{x}_i|^{N_i}$ ($N_i \geq 1$: integer). The scalar quantum potential becomes singular at the strong nodal points:

$$V_Q = -(\hbar^2/2m)\nabla^2 R/R \approx -(\hbar^2/2m)N_i^2/|\mathbf{r} - \mathbf{x}_i|^2 \quad (2)$$

This form of quantum potential leads to the formation of quantized vortices in which the velocity field attains

the magnitude: $v = N_i \hbar / (m |\mathbf{r} - \mathbf{x}_i|)$. There are analogies with a classical vortex filament, but we note that the current density remains finite: $\mathbf{J} \propto |\mathbf{r} - \mathbf{x}_i|^{2N-1}$.

3. Examples of Vortex States

Vortex formation is intimately related to the projection of the flow into a pure angular momentum state. For example, the states of a coherent 2DEG electron confined to a closed circular quantum dot include eigenstates of angular momentum $M\hbar$ states for which the flow is a pure vortex with velocity field of magnitude $v_\phi = \frac{M\hbar}{mr}$ (Fig. 1). Generalising to a cylindrical dot there are drifted angular momentum states of the form $\Psi \sim (k_z r)^M \exp[i(M\phi + k_z z - et/\hbar)]$ for which the amplitude and phase satisfy $R \sim (k_z r)^M$ and $S = M\hbar\phi + \hbar k_z z - et = \hbar \{ \text{ArcTan}[y/x] + k_z z - et/\hbar \}$. The corresponding flow has a central vortex line and the quantum trajectories (Barker 2001) are helical (see Fig. 2). It is anticipated that any obstacle to the velocity flow that may generate angular momentum will lead to vortex formation. Indeed, we have found this to be the case in the numerical solution to the time-dependent Schrödinger equation (TDSE) for flows of electrons in coupled open quantum dots in the presence of atomistic impurities. Figure 3 shows the vortex formation in the velocity field, the particle probability density in contour and landscape form for a gaussian wavepacket travelling along a 2D quantum waveguide containing an open quantum dot. This result is a frame from a sequence of solutions to the TDSE computed using a new high speed algorithm (Barker, Watling and Wilkins 2001). This example is a 2D analogue of Kelvin's smoke ring vortex experiment. In Fig. 4 we display the corresponding quantum potential and note that it is not singular (soft core) at the vortex centres. This work

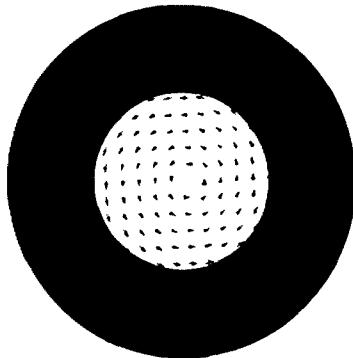


Figure 1. Velocity flow in a circular quantum dot: vortex filament.

$$v_\phi = \frac{M\hbar}{mr}$$

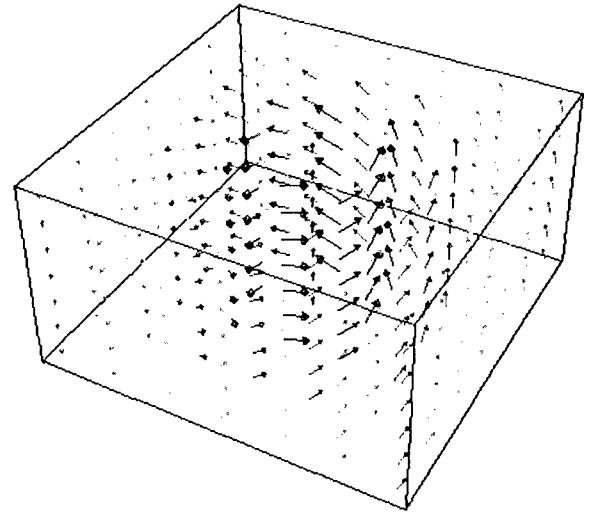


Figure 2. Helical velocity flow in idealised cylindrical quantum dot.

is part of a systematic study of vortex formation in the electron flow past obstacles (including impurities) in open quantum dot structures (Fig. 5). We note that the topological model of quantum flows developed in (Barker 2001) explains the vortex pairs in Fig. 4 in terms of the flow repulsion of classical trajectories reflecting of the exit walls of the quantum dot and which would otherwise cross. Finally we note that by adding a coupled magnetic field to the Hamiltonian leads to states in which circulation may occur due to cyclotron motion as well as orbital angular momentum induced by the geometry. For example the familiar Landau states take on a revealing form if cylindrical boundary conditions are imposed (for example in a quantum dot with perpendicular magnetic field). It is easy to show that the states and velocity flows satisfy:

$$\begin{aligned} \Psi(r, \phi, z) &= R_{NM}(r) e^{iM\phi} e^{ik_z z} \\ \epsilon &= \hbar\omega_c \left\{ \tilde{N} + \frac{1}{2}(1 + |M| - M) \right\} + p_z^2/2m \quad (3) \\ R_{NM}(r) &\sim \left(\frac{r^2}{2l^2} \right)^{|M|/2} \exp \left[-\left(\frac{r^2}{2l^2} \right)/2 \right] {}_1F_1 \\ &\times \left[-\tilde{N}, 1 + |M|, \left(\frac{r^2}{2l^2} \right) \right] \\ v_\phi &= \frac{M\hbar}{m * r} - \frac{1}{2}\omega_c r \\ \mathbf{v} &= \left\{ \frac{2M\hbar}{m * r^2} - \omega_c \right\} \mathbf{e} \quad (4) \\ \mathbf{e} &= \frac{1}{2} \{-y\mathbf{i} + x\mathbf{j}\} \end{aligned}$$

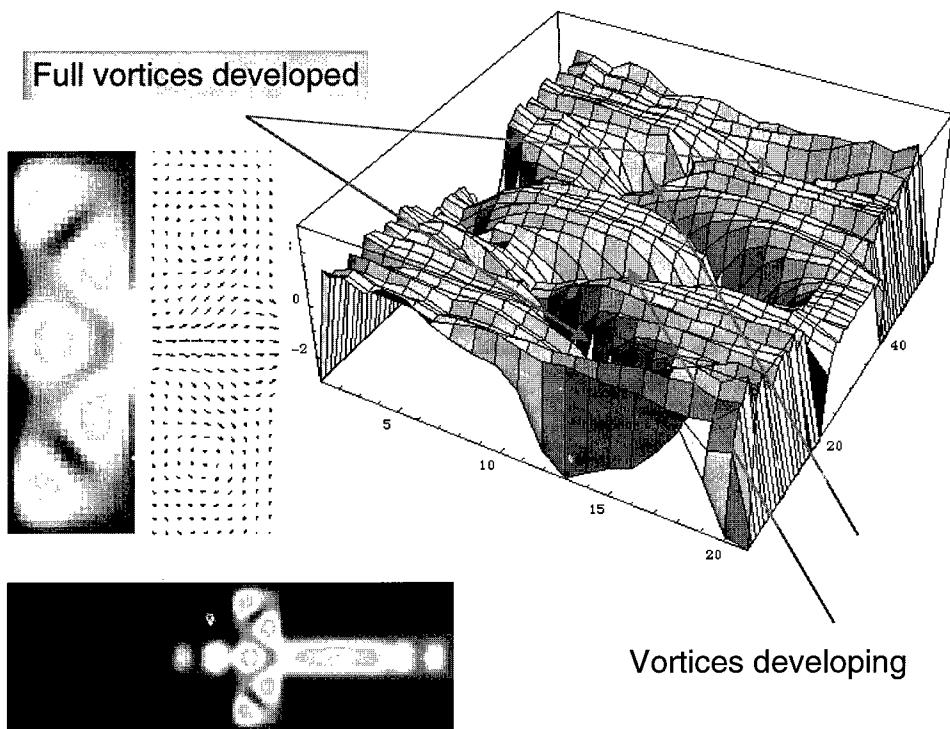


Figure 3. Particle probability density in density plot and landscape plot for a gaussian wavepacket moving through an open quantum dot showing vortex pair formation in the velocity flow (arrows).

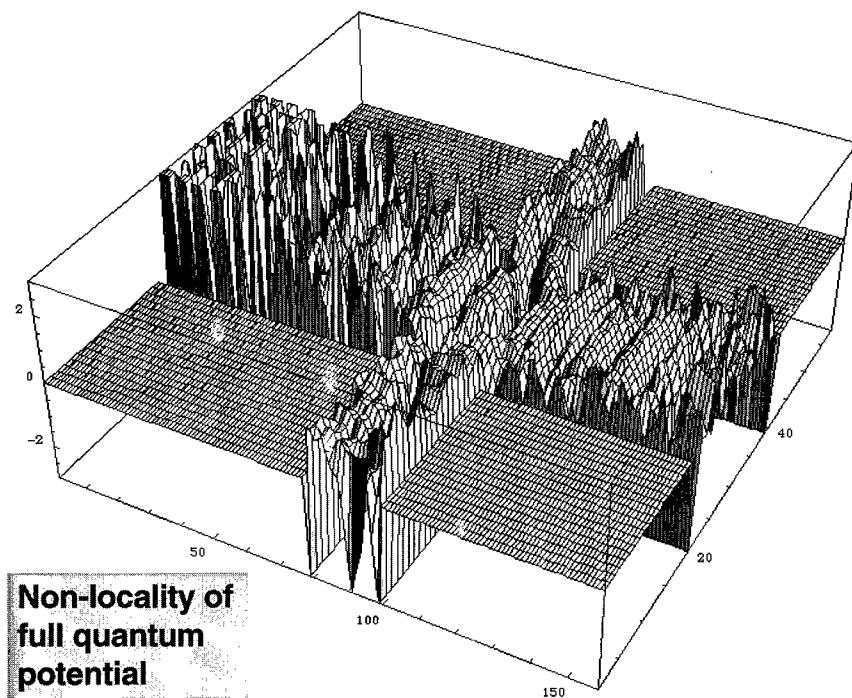


Figure 4. Quantum potential over the open quantum dot corresponding to Fig. 3.

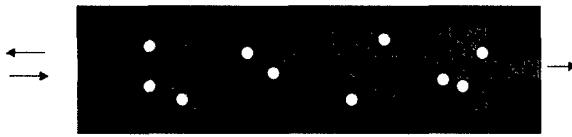


Figure 5. Schematic of coupled quantum dots with discrete impurity sites.

Here we obtain a strong nodal line through $r = 0$, and the flow comprises a quantized filamentary vortex (angular momentum $M\hbar$) and a classical vortex arising from the cyclotron motion. These flows have opposite sign (Fig. 6) and hence the velocity is zero along the stationary Landau orbit at $r = \sqrt{2(\tilde{N} + M + 1/2)}l$ where $\omega_c = eB/m$, $l = \sqrt{\hbar/m\omega_c}$.

4. Difficulties

In the above examples there is a problem with the magnitude of the velocity field which diverges on a vortex line, although the current density and particle density vanish. From a computational point of view, solving the ab initio quantum hydrodynamic (QHD) equations (which is essentially the route taken in density gradient device modelling or quantum Monte Carlo) for the velocity flow will be unstable near a vortex core. There is no such stability problem if the velocity field is deduced from a solution to the TDSE. Our numerical studies

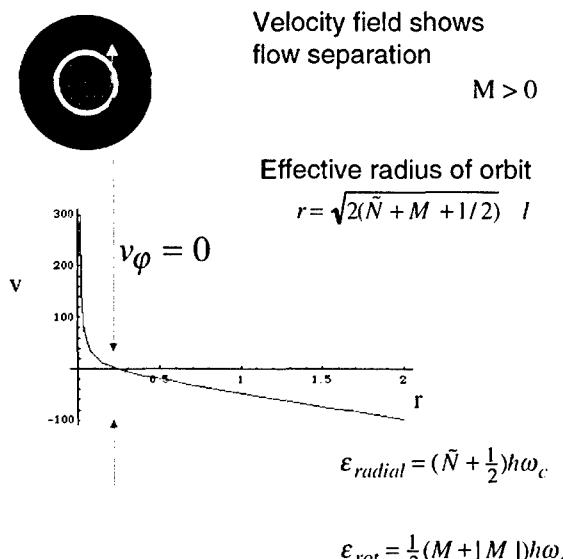


Figure 6. Velocity flow separation in electron flow in a cylindrical Landau state.

(Fig. 3) also show that vortex flows with soft cores (non-divergent velocity) occur for time-dependent flows in 2DEGs but there are still stability problems for QHD. There is also a contradiction if ab initio QHD is pursued since the equations of QHD are only valid for *irrotational flow* (as may be seen by a careful derivation from the Schrödinger theory) whereas the existence of vortices implies $\text{curl } \mathbf{v} \neq 0$.

5. The Vector Quantum Potential

To describe vortex motion self-consistently with ab initio quantum hydrodynamics we propose the introduction of a *vector* quantum potential $\mathbf{a}(\mathbf{r}, t)$ into the formalism of a fully gauge-invariant quantum hydrodynamics. This new term appears in the quantum Euler equation as a force field $\mathbf{F} = -m\mathbf{v} \times (\nabla \times \mathbf{a})$. It accounts exactly for the possibility that the velocity field $\mathbf{v} = \mathbf{J}(\mathbf{r}, t)/n(\mathbf{r}, t)$ is not everywhere irrotational: $\nabla \times \mathbf{v} \neq 0$. In the vicinity of a strong nodal line for example we find the vector quantum potential obeys:

$$\nabla \times \mathbf{a} = N_i h \int \delta(\mathbf{r} - \mathbf{x}_i(s))(d\mathbf{x}_i/ds) ds \quad (5)$$

This leads to the formation of quantized vortices with circulation $N_i h/m$. Equation (4) should be regarded as the source equation for the field \mathbf{a} , where the vortex line must be determined from a separate solution to the angular momentum density equations either from a projection of the true quantum state or by using the ab initio angular momentum density continuity equations.

A fully gauge-invariant form of Eulerian QHD is then possible with the additional inclusion of electromagnetic fields via the vector and scalar potentials \mathbf{A} and Φ .

$$m\mathbf{v} = \nabla S + \mathbf{a}_Q - e\mathbf{A} \quad (6)$$

$$\oint \{m\mathbf{v} - \mathbf{a}_Q + e\mathbf{A}\} \cdot d\mathbf{r} = nh \quad (7)$$

$$m \frac{\partial \mathbf{v}}{\partial t} + m\mathbf{v} \cdot \nabla \mathbf{v} = -\nabla[\Phi(\mathbf{x}, t) + V_Q(\mathbf{x}, t)] - m\mathbf{v} \times \nabla \times (\mathbf{a}_Q(\mathbf{x}, t) - e\mathbf{A}) \quad (8)$$

Condition (7) is the general condition for velocity circulation and permits a fully QHD picture of for example the Aharonov-Bohm effect.

These concepts may be extended to mixed-state problems based on Wigner functions, Quantum Monte

Carlo and finite temperature Quantum Hydrodynamics and will described elsewhere.

6. Discussion and Conclusions

Our studies suggest that there are strong possibilities for vortex formation in the transport of carriers in de-canano FETs due to angular momentum generation by flow through the atomistic fluctuation potential arising from the discrete impurities. QHD modelling of such devices will require a capability to describe vortex formation and destruction if it is to accurately account for fluctuation phenomena. In conclusion we have demonstrated that vortex formation corresponds

to the coupling to a novel type of vector quantum potential which leads to the quantization of velocity circulation in a similar fashion to the vector potential associated with a quantizing magnetic field.

References

- Barker J.R. 2001. New Phenomena in Mesoscopic Devices, Hawaii, Proceedings.
- Barker J.R. VLSI Design, in the press.
- Barker J.R. and Ferry D.K. 1998. Semicond.Sci.Technol. 13: A135–A139.
- Barker J.R., Ferry D.K., and Akis R. 2000. Superlatt. and Microstructures 27: 319–325.
- Barker J.R., Watling J., and Wilkins R. VLSI Design, in the press.
- Lent C.S. 1990. Appl.Phys.Lett. 57: 1678–1680.



On the Current and Density Representation of Many-Body Quantum Transport Theory

JOHN R. BARKER

Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8LT, Scotland, UK

jbarker@elec.gla.ac.uk

Abstract. The possibility of developing an extension of density functional methods but using generalised currents as coordinates is examined as a possible route for future device modelling at atomistic scales in the presence of strong many body effects.

Keywords: functional methods, semiconductor transport theory

1. Introduction

There have been great successes for density functional theory in computational chemistry and many body theory, particularly for the basic ground state structure. The success has been largely due to the minimum energy theorem for ground states, which has an analogous power to that of thermodynamics for equilibrium states. In particular for ground states it is only necessary to work with the single particle density rather than the full many-body wavefunction or density matrix. It is pertinent to investigate whether a similar formalism is available for *open* many-body systems such as the interacting inhomogeneous carrier gas in a semiconductor device. As semiconductor devices push into the 20–30 nm scale, atomistic effects, strong many body processes and significant environmental coupling suggest that an approach is required to transport and switching based on a more radical formalism than the Wigner equation or the density gradient/quantum Monte Carlo methodology. Indeed, as the possibility of novel quantum computing devices and devices based on nanotubes and biochemical structures emerge there is a need for a transport formalism that builds in many-body effects and the self-consistent electronic states. Over the last ten years there has been much advocacy of non-equilibrium thermodynamic Green function techniques, but their numerical simulation has led to serious problems of convergence and stability and in a

sense they contain too much information. At first sight quantum hydrodynamics (QHD) provides a possible minimalist approach with its focus on carrier density fields, carrier velocity fields energy density fields and so on, coupled through various continuity equations. However, in a recent study of QHD we have found that there are serious difficulties with the velocity flow picture and a better approach is to use the current densities explicitly. In this paper we therefore make a preliminary examination of a new approach to quantum transport theory which we wish to base on an old idea due to Dashen and Sharp (1968) that quantum mechanics may be described by using currents \mathbf{J} and densities ρ as coordinates rather than the usual $\{\mathbf{r}, \mathbf{p}\}$ of canonical phase-space variables. This approach is appealing because it should shed light on many-body quantum hydrodynamics where the inclusion of interactions has so far been phenomenological (relaxation time models). The overall aim would be to devise non-perturbative formalism with a simple dependence of system properties on the density and currents.

2. Difficulties with QHD

It is well-known that the equations of pure-state quantum hydrodynamics may be derived by taking the polar form for the wavefunction and separating out the real and imaginary parts of the Schrödinger equation.

This results in the continuity equation for the amplitude squared $n = R^2$ and a Hamilton Jacobi-like equation (for the phase S of the wavefunction) the gradient of which leads to an Euler-like equation for the velocity field defined by

$$\mathbf{v} = \mathbf{J}(\mathbf{r}, t)/\rho(\mathbf{r}, t) \quad (1)$$

(here the particle current density is $\mathbf{J} = \rho(\mathbf{r}, t)\nabla S(\mathbf{r}, t)/m$). There are two differences with classical hydrodynamic models: first, there appears a scalar quantum potential (from now on choosing units $\hbar = m = 1$)

$$V_Q = \left(\frac{1}{8}\right) \left\{ \left(\frac{\nabla\rho}{\rho}\right)^2 - \frac{2\nabla^2\rho}{\rho} \right\} \quad (2)$$

within the Euler equation; secondly, the resulting equations of motion are not complete, there remains an additional constraint imposed by the single-valuedness of the wavefunction leading to the quantization of velocity circulation (Barker):

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = Nh \quad (3)$$

Any spatial circuit C through which a vortex occurs in the velocity flow leads to a non-zero integer in condition (3). Vortex cores occur along the strong nodal lines (Barker) of the amplitude the wave-function, along which the phase S is indeterminate. The scalar quantum potential (2) becomes singular at the strong nodal points and within the quantized vortices the velocity field attains the magnitude: $v = N_i/|\mathbf{r} - \mathbf{x}_i|$ which diverges along the nodal line $\mathbf{r} = \mathbf{x}_i$. This non-physical result illustrates the danger with over-interpreting the velocity field and more seriously leads to numerical instabilities in solutions to the QHD equations for the velocity flow. However, the carrier density and current density remain finite at the vortex core: $\rho \approx |\mathbf{r} - \mathbf{x}_i|^{2N}$ and $\mathbf{J} \propto |\mathbf{r} - \mathbf{x}_i|^{2N-1}$ where N is a positive integer. Recently (Barker) we have shown that the constraint (3) may be incorporated into QHD via a vector quantum potential, but the fact remains that there is an intrinsic problem with a formalism based on the velocity field (1). Can we then develop a formalism based on density and currents? One approach was advocated by Dashen and Sharp (1968) for stationary quantum mechanics and extended by Pardee, Schessinger and Wright (1968) to stationary many-body problems. In the following we look at an extension of their approach from the point of view of transport theory focussing mainly on the one particle problem. The full many-body version will be discussed elsewhere.

3. The Current-Charge Density Formalism

In the Dashen-Sharp formalism (Dashen and Sharp 1968) the current and density operators are introduced in the usual way as bi-linear combinations of quantum field operators. To illustrate the ideas in a simple fashion we focus on spinless bosons in the present paper without loss of generality. The density and current density operators are then:

$$\begin{aligned} \rho(\mathbf{x}) &= \psi^+(\mathbf{x})\psi(\mathbf{x}) \\ \mathbf{J}(\mathbf{x}) &= \left(\frac{1}{2i}\right)[\psi^+(\mathbf{x})\nabla\psi(\mathbf{x}) - (\nabla\psi^+(\mathbf{x}))\psi(\mathbf{x})] \end{aligned} \quad (4)$$

with commutation relations:

$$\begin{aligned} [\rho(\mathbf{x}), \rho(\mathbf{y})] &= 0 \\ [\rho(\mathbf{x}), J_\mu(\mathbf{y})] &= -i\frac{\partial}{\partial x_\mu}[\delta(\mathbf{x} - \mathbf{y})\rho(\mathbf{x})] \\ [J_\mu(\mathbf{x}), J_\nu(\mathbf{y})] &= -i\frac{\partial}{\partial x_\nu}[\delta(\mathbf{x} - \mathbf{y})J_\mu(\mathbf{x})] \\ &\quad + i\frac{\partial}{\partial y_\mu}[\delta(\mathbf{x} - \mathbf{y})J_\nu(\mathbf{x})] \end{aligned} \quad (5)$$

The total momentum \mathbf{P} , the Hamiltonian H and the total number of particles N are given by:

$$\begin{aligned} \mathbf{P} &= \int \mathbf{J}(\mathbf{x}) d^3x \quad H = H_0 + V + U \\ H_0 &= \int H_0(\mathbf{x}) d^3x = \frac{1}{8} \int d^3x [\nabla\rho(\mathbf{x}) - 2i\mathbf{J}(\mathbf{x})] \\ &\quad \times \frac{1}{\rho(\mathbf{x})} \cdot [\nabla\rho(\mathbf{x}) + 2i\mathbf{J}(\mathbf{x})] \\ V &= \frac{1}{2} \int d^3x d^3y \rho(\mathbf{x})\rho(\mathbf{y})V(|\mathbf{x} - \mathbf{y}|) \\ U &= \frac{1}{2} \int d^3x \rho(\mathbf{x})U(\mathbf{x}) \quad N = \int d^3x \rho(\mathbf{x}) \end{aligned} \quad (6)$$

Following Sharp (1968), the key step is to introduce a functional representation of the algebra (5-6). we choose the eigenvectors of the density operator as basis states and within that space we define a wave functional: $\Psi(\rho) = \langle \rho | \Psi \rangle$. The scalar product is defined by the functional form:

$$\langle \Psi | \Phi \rangle = \int \Psi(\rho)^+ \Phi(\rho) D(\rho) \quad (7)$$

The measure D was not defined by Dashen and Sharp but the functional integration may done over the eigenvalues of the density operator which are delta functions located at the position of each particle. The action of

the current operator on these states is then represented by the functional derivative:

$$\mathbf{J} \rightarrow -i\rho(\mathbf{r})\nabla \frac{\delta}{\delta\rho(\mathbf{r})} \quad (8)$$

The energy spectrum follows from:

$$H\Psi(\rho) = E\Psi(\rho) \quad (9)$$

4. Relation to QHD

We observe from (5) that in this picture the quantum potential (2) appears naturally (note the Laplacian term in (5) integrates out). In particular we may write:

$$\begin{aligned} H_0 &= \int H_0(\mathbf{x}) d^3x = \frac{1}{8} \int d^3x [\nabla\rho(\mathbf{x}) - 2i\mathbf{J}(\mathbf{x})] \\ &\quad \times \frac{1}{\rho(\mathbf{x})} \cdot [\nabla\rho(\mathbf{x}) + 2i\mathbf{J}(\mathbf{x})] \\ H_0 &= T_0 + T_Q = \frac{1}{2} \int d^3x \mathbf{J}(\mathbf{x}) \frac{1}{\rho(\mathbf{x})} \cdot \mathbf{J}(\mathbf{x}) \\ &\quad + \frac{1}{8} \int d^3x \nabla\rho(\mathbf{x}) \frac{1}{\rho(\mathbf{x})} \cdot \nabla\rho(\mathbf{x}) \end{aligned} \quad (10)$$

where

$$\begin{aligned} T_Q[\rho] &= \frac{1}{8} \int d^3x \nabla\rho(\mathbf{x}) \frac{1}{\rho(\mathbf{x})} \cdot \nabla\rho(\mathbf{x}) \\ V_Q &= \frac{\delta T_Q[\rho]}{\delta\rho} = \left(\frac{1}{8}\right) \left\{ \left(\frac{\nabla\rho}{\rho}\right)^2 - \frac{2\nabla^2\rho}{\rho} \right\} \end{aligned} \quad (11)$$

The quantum potential of QHD is thus the density functional derivative of the density-dependent part of the kinetic energy operator. Equations (10) and (11) provide a starting point for the development of a generalised hydrodynamic picture based on continuity equations for currents and densities and with an explicit quantum potential.

5. Single Particle Picture

For a single particle we may first find the functional corresponding to particle at rest, which requires

$$\begin{aligned} H_0\Psi_0 &= 0 \Rightarrow [\nabla\rho(\mathbf{x}) + 2i\mathbf{J}(\mathbf{x})]\Psi_0(\rho) = 0 \\ &= \left(\nabla\rho(\mathbf{x}) + 2\rho \nabla \frac{\delta}{\delta\rho} \right) \Psi_0(\rho) \\ &= \left(\frac{1}{2\rho} \nabla\rho(\mathbf{x}) + \nabla \frac{\delta}{\delta\rho} \right) \Psi_0(\rho) = 0 \Rightarrow \quad (12) \end{aligned}$$

(if $\rho \neq 0$)

$$\nabla \left(\frac{1}{2} \ln \rho(\mathbf{x}) + \frac{\delta}{\delta\rho} \right) \Psi_0(\rho) = 0$$

Equation (12) has the solution (by analogy with first order differential equations):

$$\Psi_0 = \exp \left[-\frac{1}{2} \int \rho(\mathbf{x}) \ln \rho(\mathbf{x}) d^3x \right] \quad (13)$$

Similarly, using (6) we find the functional for a free particle with momentum \mathbf{p} as:

$$\Psi_0 = \exp \left[\int \left\{ i\mathbf{p} \cdot \mathbf{x} - \frac{1}{2} \ln \rho(\mathbf{x}) \right\} \rho(\mathbf{x}) d^3x \right] \quad (14)$$

Guided by these results, we examine a single particle in a potential $U(\mathbf{x})$ and look for a functional in the form

$$\Psi(\rho) = \exp \left[\int \rho(\mathbf{x}) \left\{ g(\mathbf{x}) - \frac{1}{2} \ln \rho(\mathbf{x}) \right\} d^3x \right] \quad (15)$$

From the eigenvalue equation

$$H\Psi(\rho) = E\Psi(\rho) = E \int \rho(x) d^3x \Psi(\rho) \quad (16)$$

where we have used the integral over the density is unity for one particle, we obtain:

$$\int d^3x \left\{ \frac{1}{2} \nabla^2 g + \frac{1}{2} (\nabla g)^2 + U - E \right\} \rho = 0 \quad (17)$$

The density is arbitrary so the integrand of (17) must vanish and using the transformation

$$g = -\ln \varphi \quad (18)$$

we find φ is a solution to the Schrödinger equation:

$$\left\{ -\frac{1}{2} \nabla^2 + U \right\} \varphi = E\varphi \quad (19)$$

These results were first obtained by Pardee, Schessinger and Wright (1968) using the Dashen-Sharp formalism (Dashen and Sharp 1968).

6. The Gutter Potential

The gutter potential has been widely used in transport modelling; it describes a quantum waveguide in a 2DEG in the form:

$$U(x, y) = \frac{1}{2} w^2 y^2 - Fx \quad (20)$$

thus the potential is confining in the y -direction and is open in the x -direction. In the case $F=0$ we have a confining harmonic oscillator potential and the corresponding density functional is found to be:

$$\Psi_m = \exp \left[\int \frac{1}{2} \left\{ 2ipx + \ln [H_m^2(w^{1/2}y) - \ln \rho(x, y)] \right\} \rho(x, y) dx dy \right] \quad (21)$$

where H_m is the m th order Hermite Polynomial. In the case $F \neq 0$ the term in ipx in (21) is replaced by $\ln((2\pi^{-3/2}F^{-1/2})^{1/3}Ai(-x))$ where Ai is the Airy function.

The expectation values for the current and density operators in the state $\Psi(\rho)$ give the usual density and current obtained in orthodox quantum mechanics.

For N particles in the ground state of the confining potential ($F=0$) the density functional becomes:

$$\Psi = \exp \left[\int \frac{1}{2} \left\{ w y^2 - \ln \rho(x, y) \right\} \rho(x, y) dx dy \right] \quad (22)$$

Important correlation functions such as the two-particle correlation function G may be computed easily. For example G for the ground state is given by $\langle \rho(y)\rho(y') \rangle$ and it may be determined from the projection of the operator equation

$$[\rho(y), \mathbf{J}(y')] = -i \nabla [\delta(y - y')\rho(y)]$$

on the ground state and using the hermiticity of the anti-commutator of $\mathbf{J}(y)$ and $\rho(y')$. The resulting differential equation has the well-known solution:

$$G = \langle \rho(y)\rho(y') \rangle = \langle \rho(y) \rangle \delta(y - y') + (1 - 1/N) \langle \rho(y) \rangle \langle \rho(y') \rangle.$$

7. Conclusions

A time-dependent formalism may be constructed by replacing the stationary Schrödinger equation (9) by the time-dependent form:

$$i \frac{\partial}{\partial t} |\Psi\rangle = H |\Psi\rangle. \quad (23)$$

The densities and currents are now defined over the full space-time domain. In particular, it is found that the generalised continuity equation holds (by forming the commutator $[H, \rho]$)

$$\nabla \cdot (\mathbf{J}) + \frac{\partial}{\partial t} \langle \rho \rangle = 0. \quad (24)$$

In the non-interacting case we may derive the time-dependent generalisation of (15)

$$\Psi(\rho, t) = \exp \left[\int \rho(\mathbf{x}) \left\{ g(\mathbf{x}, t) - \frac{1}{2} \ln \rho(\mathbf{x}) \right\} d^3 \mathbf{r} \right] \quad (25)$$

where $g(\mathbf{x}, t) = -\ln \varphi(\mathbf{x}, t)$ is related to the solution $\varphi(\mathbf{x}, t)$ of the time-dependent Schrödinger equation $(-\frac{1}{2}\nabla^2 + U)\varphi = i(\partial/\partial t)\varphi$. This brief overview of the formulation of the current and density formalism shows that the quantum potential occurs naturally in this picture, there are no divergence issues with using the currents and densities unlike conventional QHD. In future reports we aim to explore the extension of this picture to a generalised hydrodynamic picture of transport based on non-perturbative functional methods.

References

- Barker J.R. these proceedings.
- Dashen R.F. and Sharp D.H. 1968. Phys. Rev. 165: 1857.
- Pardee W.J., Schessinger L., and Wright J. 1968. Phys. Rev. 175: 2140.
- Sharp D.H. 1968. Phys. Rev. 165: 1867.



A Space Dependent Wigner Equation Including Phonon Interaction

M. NEDJALKOV, H. KOSINA, R. KOSIK AND S. SELBERHERR

Institute for Microelectronics, TU Vienna, Gusshausstrasse 27–29, A-1040 Vienna, Austria

Nedjalkov@iue.tuwien.ac.at

Abstract. We present a kinetic equation which is obtained after a hierarchy of approximations from the generalized Wigner function equation which accounts for interaction with phonons. The equation treats the coherent part of the transport imposed by the nanostructure potential at a rigorous quantum level. It is general enough to account for the quantum effects in the dissipative part of the transport due to the electron-phonon interaction. Numerical experiments demonstrate the effects of collisional broadening, retardation and the intra-collisional field effect. The obtained equation can be regarded as a generalization of the Levinson equation for space dependence. An analysis shows that the equation is nonlocal in the real space. This quantum effect is due to the correlation between the interaction process and the space component of the Wigner path.

Keywords: Wigner function, nanostructure, quantum electron-phonon interaction, Monte Carlo method

Introduction

The quantum transport in far from equilibrium conditions is determined not only by the nanoscale of the device potential, but also by dissipative processes due to interaction with phonons. Usually the boundary conditions are given by electrons in traveling states entering into a nanodevice from the leads. If only the coherent part of the transport is considered, these states remain isolated from the notch states, which exist at the lower energy regions of the device potential. In this case unphysical simulation results can be obtained (Frenksley 1990). Thus dissipative processes which are due to interaction with phonons must be taken into account. The electron-phonon interaction links the traveling and the notch states and correctly redistributes the electrons into the device. It has been shown that the electron phonon interaction greatly affects the device parameters of the resonant tunneling diodes (Zhao *et al.* 2001). While the theoretical and numerical aspects of the application of the coherent Wigner equation are well established, the inclusion of the electron-phonon interaction is still under investigation.

Approximations

A rigorous inclusion of the phonon interaction is provided by the generalized Wigner function (WF) (Bordone *et al.* 1999) $f_w(\mathbf{r}, \mathbf{p}, \{n\}, \{m\}, t)$ which along with the electron coordinates \mathbf{r}, \mathbf{p} depends also on the phonon coordinates $\{n\} = \{n_1, \dots, n_q, \dots\}$ with n_q being the number of phonons in mode \mathbf{q} . Of interest is the reduced WF $f_w(\mathbf{r}, \mathbf{p}, t)$, which is obtained by taking the trace of the generalized WF over the phonon system and thus depends only on the electron coordinates. An exact equation for the reduced WF can not be obtained from the generalized Wigner equation, since the trace operation does not commute with the electron-phonon interaction Hamiltonian.

The task is to obtain from the generalized Wigner equation a closed equation for the reduced WF. The approximations include a weak scattering limit in the phonon interaction, assumption of an equilibrium phonon system, mean phonon number approximation, and an effective field in the scattering-Wigner potential correlation.

The generalized Wigner equation couples an element $f_w(\dots, \{n\}, \{m\}, t)$ to four neighborhood elements

given by $f_w(\dots, \{n\} \pm 1_{\mathbf{q}}, \{m\}, t)$, $f_w(\dots, \{n\}, \{m\} \pm 1_{\mathbf{q}}, t)$ for any phonon mode \mathbf{q} . The equations for the four neighboring elements involve elements which are secondary neighbors with respect to the $(\{n\}, \{m\})$ element. In this way the diagonal elements, involved in the trace operation are linked to all off-diagonal elements. As a first approximation we consider the weak scattering limit, which neglects all links to the elements placed outside the nearest off-diagonals. This assumption ignores higher order electron phonon interactions.

The evolution process begins with an initially decoupled electron-phonon system and involves transitions between the diagonal and the first off-diagonal elements. The next approximation is to replace the occupation numbers $n_{\mathbf{q}}$ involved in the transitions with the equilibrium phonon number $n(\mathbf{q})$: This is done by performing the trace operation at the consecutive time steps of the evolution. With this it is assumed that the phonons stay in equilibrium during the evolution (phonon bath). This allows to perform the trace operation and to obtain a closed equation set for the reduced WF. The set consists of a main equation for the reduced WF coupled to two auxiliary equations. The latter arise from the first off-diagonal terms of the generalized WF and describe the electron-phonon interaction. While the equation for the reduced WF is real, the two auxiliary equations are complex and mutually conjugated. The formal solution of the auxiliary equations is given by the Neumann series, which can be substituted into the main equation.

The implicit inclusion of the Neumann expansions in the main equation is rather inconvenient and we look for an approximation where the two auxiliary equations can be solved explicitly. If the potential term in the two auxiliary equations is approximated by the mean homogeneous electric field \mathbf{E} throughout the device (mean field approximation), the solution to the two auxiliary equations can be explicitly expressed in terms of the reduced WF. This approximation concerns only the phonon interaction, while the potential term in the equation for the reduced WF is treated exactly. A single equation for the reduced WF is obtained.

$$\begin{aligned} f_w(\mathbf{r}, \mathbf{p}, t) &= f_0(\mathbf{r}_{(p,0)}, \mathbf{p}_{(0)}, 0) + \int_0^t d\tau' \int d\mathbf{p}' V'_w \\ &\quad \times (\mathbf{r}_{(p,t')}, \mathbf{p}' - \mathbf{p}_{(t')}) f_w(\mathbf{r}_{(p,t')}, \mathbf{p}', t') \\ &\quad + 2 \int_0^t dt' \int_0^{t'} dt'' \sum_{\mathbf{q}'} F^2(\mathbf{q}') \end{aligned}$$

$$\begin{aligned} &\times \cos \left(\int_{t''}^{t'} d\tau \frac{1}{\hbar} (\epsilon(\mathbf{p}_{(\tau)}) - \epsilon(\mathbf{p}_{(\tau)} - \hbar\mathbf{q}') - \hbar\omega_{\mathbf{q}'}) \right) \\ &\times \left\{ n(\mathbf{q}') f_w \left(\mathbf{r}_{(p,t'')} + \frac{\hbar\mathbf{q}'}{2m}(t' - t''), \mathbf{p}_{(t'')} - \hbar\mathbf{q}', t'' \right) \right. \\ &\quad \left. - (n(\mathbf{q}') + 1) f_w \left(\mathbf{r}_{(p,t'')} + \frac{\hbar\mathbf{q}'}{2m}(t' - t''), \mathbf{p}_{(t'')}, t'' \right) \right\} \\ &- 2 \int_0^t dt' \int_0^{t'} dt'' \sum_{\mathbf{q}'} F^2(\mathbf{q}') \\ &\times \cos \left(\int_{t''}^{t'} d\tau \frac{1}{\hbar} (\epsilon(\mathbf{p}_{(\tau)}) - \epsilon(\mathbf{p}_{(\tau)} + \hbar\mathbf{q}') + \hbar\omega_{\mathbf{q}'}) \right) \\ &\times \left\{ n(\mathbf{q}') f_w \left(\mathbf{r}_{(p,t'')} - \frac{\hbar\mathbf{q}'}{2m}(t' - t''), \mathbf{p}_{(t'')}, t'' \right) \right. \\ &\quad \left. - (n(\mathbf{q}') + 1) f_w \left(\mathbf{r}_{(p,t'')} - \frac{\hbar\mathbf{q}'}{2m}(t' - t''), \right. \right. \\ &\quad \left. \left. \mathbf{p}_{(t'')} + \hbar\mathbf{q}', t'' \right) \right\} \end{aligned} \quad (1)$$

Here V'_w is obtained by the Wigner transform of the device potential V corrected by the potential of the homogeneous field \mathbf{E} . The rest of the notations will be explained below.

Analysis of the Equation

The reduced WF is expressed as a sum of contributions coming from the initial distribution, the interaction of the electron with the device potential and the electron-phonon interaction.

The contributions from the first two terms to the value of $f_w(\mathbf{r}, \mathbf{p}, t)$ occur on the Newton trajectory $(\mathbf{r}_{(p,t')}, \mathbf{p}_{(t')})$ initialized by \mathbf{r}, \mathbf{p} at time t . The initial condition f_0 evolves on this trajectory and adds to f_w its value at point $(\mathbf{r}_{(p,0)}, \mathbf{p}_{(0)})$.

The term from the potential provides information to $f_w(t)$ from $f_w(t')$ at previous times $t' \in (0, t)$. This information is nonlocal in the momentum part of the phase space, but it is local in the real space part of the trajectory $\mathbf{r}_{(p,t')}, t' \in (t, 0)$. The contribution of this term can be evaluated from the knowledge of $f_w(\mathbf{r}_{(p,t')}, \mathbf{p}', t')$ at the past of the evolution defined on the real space part of the trajectory.

A novel effect arises due to the correlation between the phonon momentum $\hbar\mathbf{q}'$ and the space component of the trajectory in the scattering terms. At the beginning of the scattering, the real trajectory is shifted by $\frac{\hbar\mathbf{q}/2}{m}(t' - t'')$. The interaction proceeds in two steps, e.g. for the terms in the first curly brackets: The first half

of a phonon momentum is absorbed (emitted) at t'' . At t' the second half is absorbed—real absorption, or the first half is absorbed back (virtual emission). In both cases the position at t' is just the right one, $\mathbf{r}_{(\mathbf{p}, t')}, \mathbf{p}_{(t')}$, which evolves to \mathbf{r}, \mathbf{p} at t . The term related to the last curly brackets is analyzed in similar way. In contrast to the Wigner equation without phonon interaction, the obtained equation becomes nonlocal in the real space.

The classical limit $\hbar \rightarrow 0$ in the phonon interaction leads to a Wigner equation with a Boltzmann scattering term. For a bulk semiconductor with an applied electric field E the equation resembles the Levinson equation (Rammer 1991), or equivalently the Barker-Ferry equation without damping of the electron lifetime.

Simulation Results

We investigate equation for quantum effects which are purely due to the electron-phonon interaction. Equation (1) is written for a bulk semiconductor in presence of an applied electric field. Cylindrical coordinates (r, k, ϕ) with r chosen normal to the field direction are used in the wave vector space. A transformation is used which shifts the coordinate system in time with the electric field. To solve (1) a randomized backward Monte Carlo algorithm is applied (Gurov and Whitlock 2001).

Simulation results for *GaAs* with a PO phonon with constant energy $\hbar\omega$ are presented. The initial condition is a sharp Gaussian function of the energy. A very low temperature, where the physical system has a transparent semiclassical behavior is assumed. The solutions are obtained on cut lines parallel to the field, $(k > 0, r = 0)$, opposite to the field, $(k < 0, r = 0)$ and normal to the field, $(k = 0, r > 0)$.

Collisional Broadening and Retardation

The effects of collisional broadening and retardation exist already at zero electric field. Figures 1–3 present snapshots of the evolution of the semiclassical and quantum solutions $|k|f(0, |k|, t)$ for times 100 fs, 200 fs and 300 fs as a function of $|k|^2$. The quantity $|k|^2$ is proportional to the electron energy in units 10^{14} m^{-2} . Semiclassical electrons can only emit phonons and loose energy equal to a multiple of the phonon energy $\hbar\omega$. They evolve according to a distribution, patterned by replicas of the initial condition shifted towards low energies.

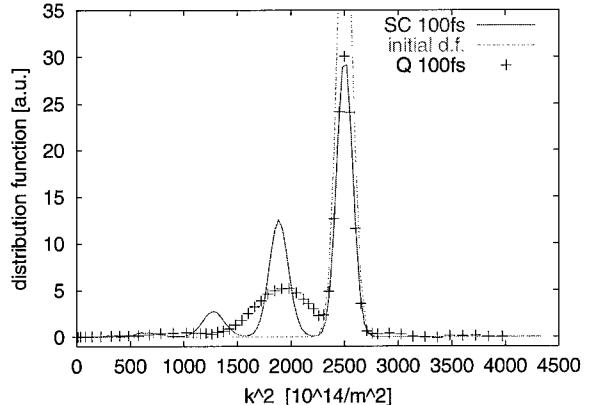


Figure 1. Initial distribution function (initial d.f.), semiclassical (SC) and quantum (Q) solutions $kf(0, k, t)$ for 100 fs evolution time at zero electric field.

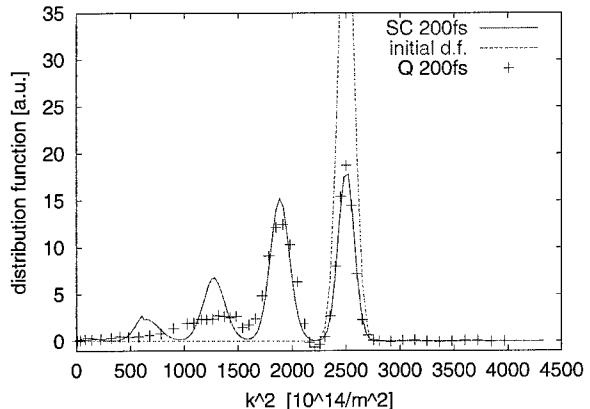


Figure 2. Initial distribution function (initial d.f.), semiclassical (SC) and quantum (Q) solutions $kf(0, k, t)$ for 200 fs evolution time at zero electric field.

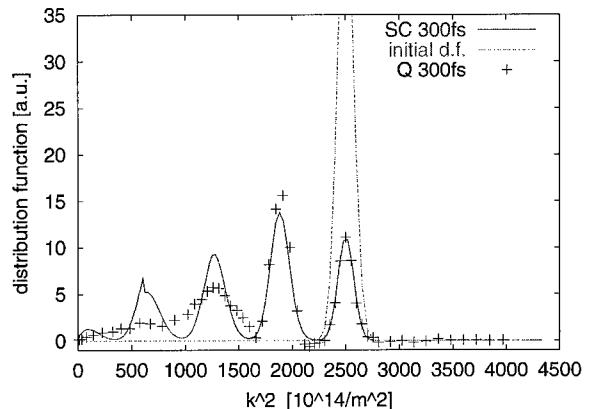


Figure 3. Initial distribution function (initial d.f.), semiclassical (SC) and quantum (Q) solutions $kf(0, k, t)$ for 300 fs evolution time at zero electric field.

The electrons cannot appear in the region above the initial distribution. The quantum solutions demonstrate two effects of deviation from the semiclassical behavior. There is a retardation in the build up of the remote peaks with respect to the initial condition peaks. The replicas are broadened and the broadening increases with the distance to the initial peak. The broadening is due to the lack of energy conservation in the interaction. At low evolution times the cosine function in (1) weakly depends on the phase space variables. With the increase of the time, the cosine term becomes a sharper function of these variables and in the long time limit tends to the semiclassical delta function. Accordingly the first replica of the 100 fs is broadened. The quantum solution resembles the main peak and the first replica of the semiclassical solution after 300 fs evolution time while the remote replicas remain broadened. The retardation of the quantum solutions is associated with the memory character of the equation. The two time integrals in (1) lead to a delay of the build up of the replicas as compared to the single time integral in the Boltzmann case.

Intra-Collisional Field Effect

Figure 4 compares the 200 fs solutions as a function of $k < 0$ for different positive values of the field. The first replica peaks are shifted to the left by the increasing electric field. The numerical solution in the semiclassically forbidden region, above the initial condition, demonstrates enhancement of the electron population with the growth of the field.

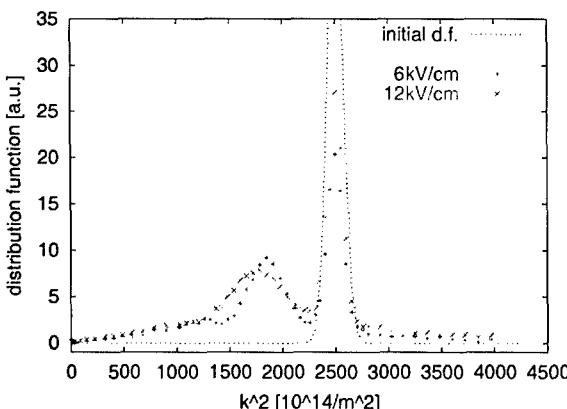


Figure 4. Solutions $|k|f(0, k, t)$, at negative k values, and evolution time 200 fs. The electric field is 0, 6 kV/cm, and 12 kV/cm.

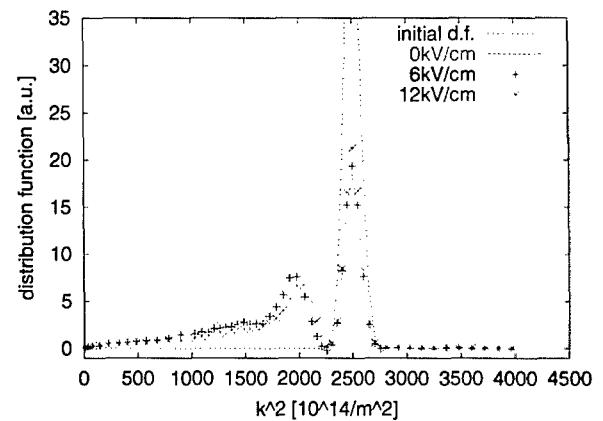


Figure 5. Solutions $kf(0, k, t)$ for positive k values and evolution time 200 fs. The electric field is 0, 6 kV/cm, and 12 kV/cm.

For states below the initial condition the energy of the field is added to the phonon energy. Accordingly the solution behaves as in presence of a phonon with energy higher than $\hbar\omega$; the distance between the first replica and the initial condition increases. For states above the initial condition the energy of the field reduces the phonon energy and thus the electron population in the vicinity of the initial condition increases. Just the opposite effects appear in the region of positive k values. This is demonstrated in Fig. 5. The peaks of the first replica are shifted to the right and there is no enhancement of the electron population above the initial condition. The field has a pronounced effect on the broadening and retardation of the solutions: A comparison of the first replicas and the main peaks under the initial condition on Figs. 4 and 5 show that the field influences the effects of collisional broadening and the retardation.

Conclusion

Starting from a full quantum mechanical model we have identified the physical assumptions necessary to derive an approximate but closed model for the reduced Wigner function. The obtained equation can be regarded as a generalization of the Levinson equation that includes the real space dependence. It is shown that the finite duration of the phonon interaction gives rise to a space non-locality of the quantum transport process. Quantum effects in electron phonon interaction have been demonstrated numerically. Observed are collisional broadening, retardation and the intra-collisional field effect.

Acknowledgment

This work has been supported by the IST program, project NANOTCAD, IST-1999-10828, and the “Christian Doppler Forschungsgesellschaft”, Vienna, Austria.

References

- Bordone P., Pascoli M., Brunetti R., Bertoni A., and Jacoboni C. 1999. Quantum transport of electrons in open nanostructures with the Wigner function formalism. *Physical Review B* 59(4): 3060–3069.
- Frensley W.R. 1990. Boundary conditions for open quantum systems driven far from equilibrium. *Reviews of Modern Physics* 62(3): 745–791.
- Gurov T. and Whitlock P. 2001. Statistical algorithms for simulation of electron quantum kinetics in semiconductors—Part I. In: *Proceedings of the 3rd International Conference on Large-Scale Scientific Computations, 2001, LNCS*, Springer.
- Rammer J. 1991. Quantum transport theory of electrons in solids: A single-particle approach. *Reviews of Modern Physics* 63(4): 781–817.
- Zhao P., Cui H.L., Woolard D.L., Jensen K.L., and Buot F.A. 2001. Equivalent circuit parameters of resonant tunneling diodes extracted from self-consistent Wigner-Poisson simulation. *IEEE Transactions on Electron Devices* 48(4): 614–627.



RTD Relaxation Oscillations, the Time Dependent Wigner Equation and Phase Noise

H.L. GRUBIN AND R.C. BUGGELN

Scientific Research Associates, Inc., Glastonbury, CT 06033, USA

hal@sraassoc.com

Abstract. Wigner simulations of resonant tunneling diode (RTD) self-excited oscillations are discussed with respect to the upper frequency limit of operation and their sensitivity to large scale perturbations. These studies offer the most practical assessment of phase noise, response times of RTDs and of the coupling of quantum well space charge to its environment.

Keywords:

Introduction

Negative differential conductivity devices sustain self-excited relaxation oscillations that arise from nonlinearities. These nonlinearities can not be regarded as small since they control the operating level of the oscillator. Phase noise exists in self excited oscillator because the latter has no time-reference. A solution to the oscillator equations that is shifted in time is still a solution. Noise can induce a time-shift in the solution, and this time-shift looks like a phase change in the signal (hence the term “phase noise”).

For a suitable set of parameters the RTD exhibits negative differential conductance and it too operates as a self excited oscillator (Verghese, Parker and Brown 1998). The properties of the RTD self excited oscillator scale much the same way as a van der Pol oscillator, so long as the device sustains negative conductance. But in general negative conductance which is a dynamic effect disappears at sufficiently high frequency. The RTD also exhibits phase noise, but the response of the system is slower than that of a self sustained oscillator with a defined region of negative conductance because the response of the carriers to any perturbation occurs over a finite period of time.

To illustrate these features the RTD was incorporated into the Fig. 1 circuit (Grubin and Buggeln to be published). In the simulation we replaced the diode

by a time dependent Wigner-Poisson algorithm that included a device capacitance, and placed that combination in parallel with an external capacitor, all in series with the other elements. When the noise calculations were performed a parallel current source was introduced into the RTD circuit.

The external circuit was treated as a boundary condition to the Wigner equation. The boundary circuit equations are:

$$\begin{aligned} \frac{dv_D}{dt} &= \frac{2\pi Z_0}{R_D} \{ i(t) - i_D(t) \}, \\ \frac{di}{dt} &= \frac{R_D}{2\pi Z_0} \left\{ v_{APPLIED}(t) - v_D(t) + i(t) \frac{R_{LOAD}}{R_D} \right\} \end{aligned} \quad (1)$$

The terms in Eq. (1) are dimensionless. The normalized quantities are obtained as follows. From the Wigner function and the calculated DC current voltage relation $I_D(V_D)$, we identify the current I_P at the NDC threshold potential energy V_P . From this $v_D = V_D/V_P$ and $i = I/I_P$. We also identify a device resistance $R_D = V_P/I_P$, a circuit impedance $Z_0 = \sqrt{L/C_D}$, a circuit period $T_{ref} = 2\pi\sqrt{LC_D}$, and the dimensionless time $t = \tau/T_{ref}$. It is important to note that for a given applied bias, device, load resistance and circuit impedance, the boundary conditions scale with time providing the device current scales with time. Our

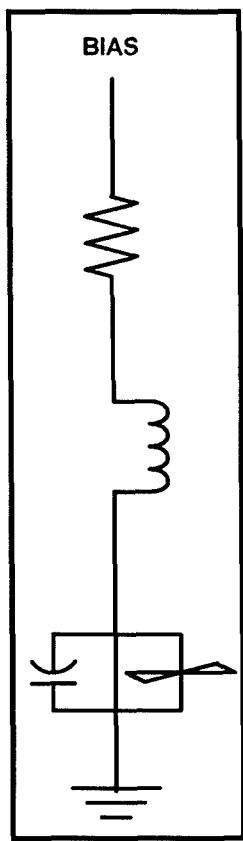


Figure 1. The circuit used for self-excited oscillations.

simulations with the Wigner function indicate that as long as the device sustains sufficient negative conductance over a cycle, the current versus time profiles are approximately independent of frequency.

The Oscillatory Characteristics and the Space Charge Profiles

Figures 2 and 3 (from Grubin and Buggeln (to be published)) illustrate the self-excited oscillation, the dynamic current voltage relation and the space charge profiles. Starting from a steady state equilibrium solution a step change in bias is introduced. Because of the presence of storage elements such as inductors and capacitors the change in potential energy across the device is gradual as displayed in Fig. 2 (dashed line). For this transient we see repetitive oscillations settling in after the first two cycles. We also show the particle current through the device (solid line). We see that after about ten ps, the potential energy across the device reaches a value of approximately 280 meV, which is the

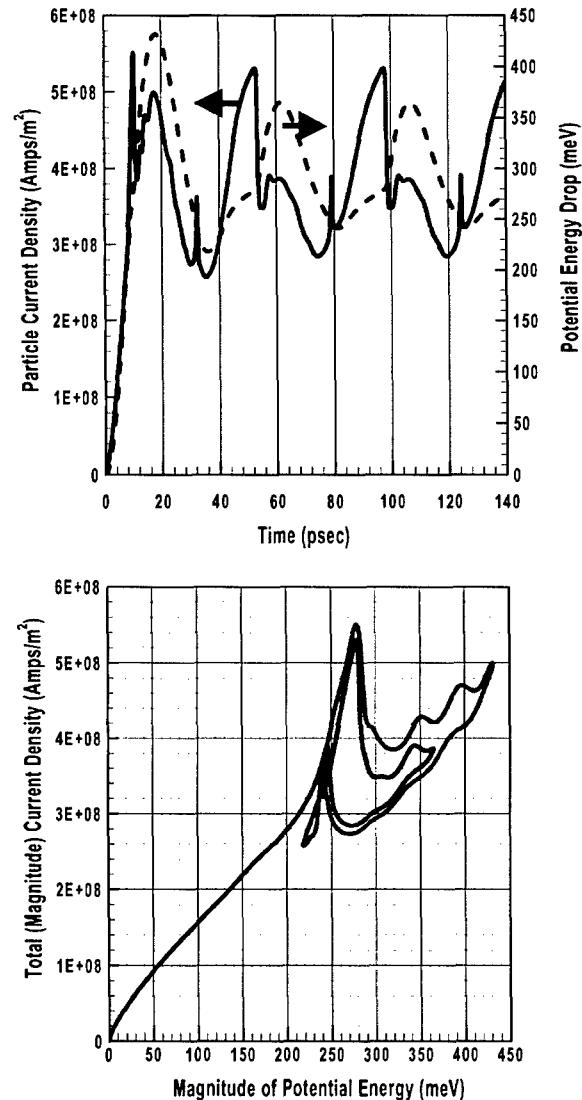


Figure 2. (Top) Transient particle current and potential energy drop for a 22 GHz oscillation. (Bottom) Several cycles of the dynamic IV for the oscillation.

threshold for NDC. There is a drop in current and a further increase in the potential energy across the device accompanied by an increased (albeit oscillatory) particle current. After passing its peak, a decreasing voltage is accompanied by a decreasing current whose values that are significantly below those accompanying the increasing voltage. This voltage decrease continues until the potential energy passes somewhat below the original NDC threshold, where there is a sudden increase in current, followed by a subsequent current decrease until a minimum is reached. The oscillation settles into a period of ~ 46 ps, for a frequency near 22 GHz. The

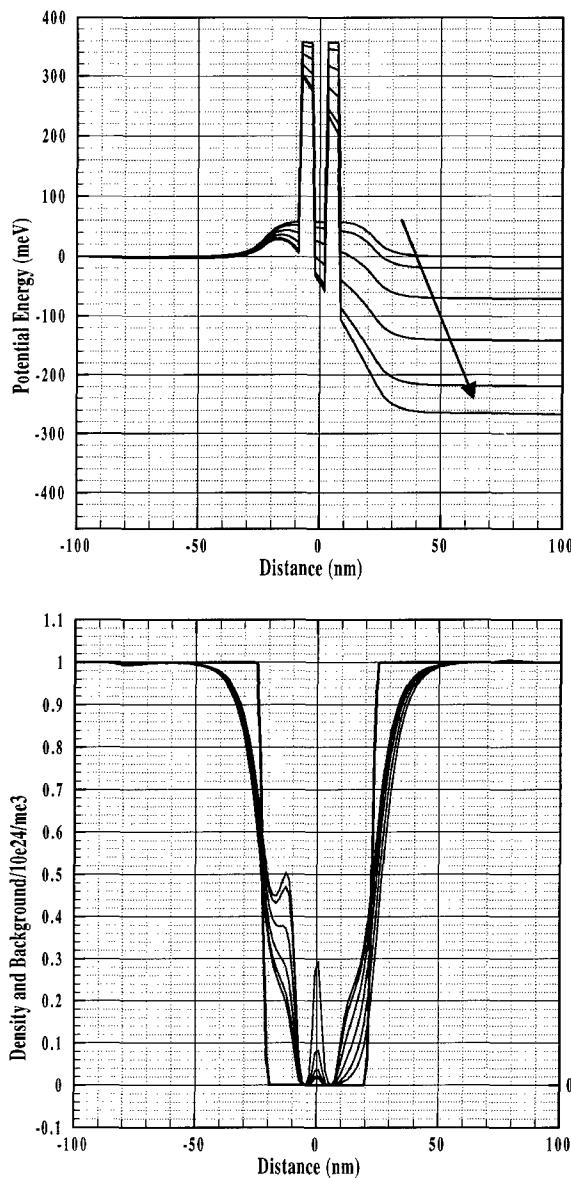


Figure 3. (Top) Potential energy and (bottom) space charge profiles for the first ~ 8 ps of the transient.

dynamic current voltage relation, which is obtained by eliminating time from the current-time and voltage-time profiles is also shown in Fig. 2, and displays the hysteresis described in the above paragraph.

The interesting feature of the Fig. 2 oscillation is that it essentially maintains this form up to about 120 GHz. There are modifications in detail, the maximum and minimum values of particle current and voltage are altered, and the NDC region weakens, with the latter feature being responsible for the cessation of oscillations.

Figure 3 displays the space charge and potential energy profiles during the first cycle of the oscillation just before NDC threshold. The features to observe are the increasing charge in the quantum well with increasing bias, as well as the formation of charge on the emitter side of the first barrier. Accompanying this is enhanced depletion on the collector side of the second barrier, satisfying the condition of global charge neutrality. Also note that as the potential energy increases, but prior to the NDC threshold, there are significant changes in the value of the potential energy of the quantum well. However, immediately prior to the current drop-back when there is significant charge accumulation in the quantum well, the voltage change within the quantum well is small compared to that across the second barrier. Indeed just prior to threshold, most of the voltage change occurs across the collector side of the structure. (The arrow in Fig. 3(a) denotes the change in potential energy as a function of bias. That in Fig. 3(b) shows the increased emitter charge accumulation with increased bias, as well as the movement of the collector charge depletion region. The quantum well charge continues to increase with bias change.) The details of Figs. 2 and 3 are discussed more fully in Grubin and Buggeln (to be published).

Effect of Perturbations on the Phase of the RTD Self-Excited Oscillation

To initiate the RTD noise study we force a change in the particle current at two instants of time, while the Wigner simulation was running. The calculation was performed for a device with the same parameters as that of Figs. 2 and 3 with the exception that the sustained oscillation occurred at 113.7 GHz. See Fig. 4. The fluctuations are indicated by the arrows.

Two important features should be noticed. First, the oscillation recovered from the perturbation within one cycle, and second there is a shift in the period. The shift in the period is the origin on the phase noise. The magnitude of the shift depends upon the duration of the fluctuation. Here the duration of the fluctuation was a substantial fraction of the oscillatory period. It also depends upon the original placement of the fluctuation.

We have also introduced fluctuations by introducing temporal variations in such quantities as the phenomenological relaxation time. Such a fluctuation might represent a temporal change in the principle type of scattering event. In each case, for the percent changes introduced the self-excited oscillation was restored.

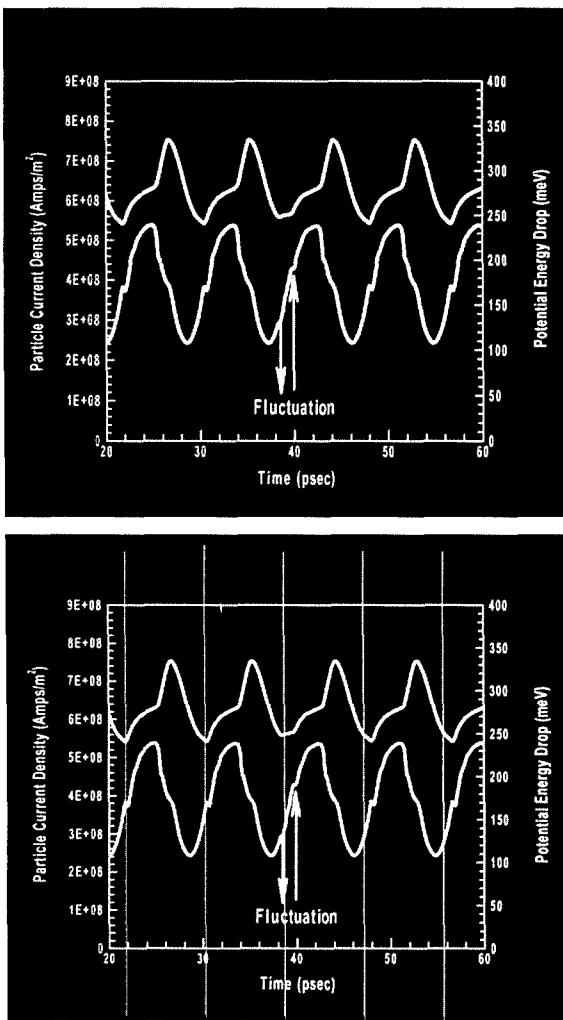


Figure 4. Top: Current (grey) and voltage (white)-time profiles prior to and after perturbation. Bottom: Same as top with period markers.

The Physical Model Used in the Simulations

The basic quantum transport equation is the Wigner equation (Wigner 1932):

$$0 = \frac{\partial f_w(\mathbf{k}, x)}{\partial t} + \frac{\hbar k_x}{m} \frac{\partial f_w(\mathbf{k}, x)}{\partial x} + \left(\frac{\partial f_w(\mathbf{k}, x)}{\partial t} \right)_{DISSIPATION} - \frac{1}{\pi \hbar} \lim_{L \rightarrow \infty} \int_{-L}^{+L} dy \times \begin{bmatrix} V(x+y) \\ -V(x-y) \end{bmatrix} \int dk'_x f_w(k'_x, k_y, k_z, y) \times \sin[2(k'_x - k_x)y] \quad (2)$$

The potential energy in Eq. (2) consists of two contributions, the barrier/well configuration (single or multiple) and the potential energy arising from Poisson's equation. In the simulations discussed here the contributions from Poisson's equation were treated classically as a term $\nabla_x V_{POISSON} \cdot \nabla_k f_w(\mathbf{k}, x)$. The barriers were square permitting an analytical integration of the Wigner integral, which was used in all of our studies (see Grubin and Buggeln to be published).

What about dissipation? The tack taken here is to relax the Wigner function, with:

$$\left(\frac{\partial f_w(\mathbf{k}, x)}{\partial t} \right)_{DISSIPATION} = -\frac{f_w(\mathbf{k}, x) - f_0(\mathbf{k}, x)}{\tau(x)} \quad (3)$$

The relaxation time approximation in the above form leads to source and sink terms in the continuity equation. To avoid sink terms others have multiplied the equilibrium distribution function by the ratio of the non-equilibrium carrier density to the equilibrium carrier density. We have done both.

The question of interest is what is $f_0(\mathbf{k}, x)$? The form of the equilibrium distribution function is dependent upon the model used to connect current at the open boundary and must represent the spatially dependent distribution associated with barriers, scattering, self-consistency, and the external circuit. The boundary condition used here sets the normal derivative, with respect to position, of the distribution function to zero. This provides the requisite zero current conditions. Further, to enhance the possibilities of flat-band open boundary conditions the relaxation time in the vicinity of the boundaries was set at least an order of magnitude smaller than elsewhere in the devices.

In performing the simulations we break the device into a classical and quantum region. The bounding reservoir region is treated classically, with the central region representing the quantum mechanical region. Within this framework the Wigner integral is multiplied by a modulating function that is equal to unity within the 'quantum region' and zero elsewhere. For calculations discussed here the device length was 200 nm, the quantum region was at least 120 nm long and included the cladding regions.

Conclusions

The study indicates that the RTD can operate as a self excited oscillator and that it can recover from perturbations in the current. These perturbations introduce

changes in phase, which are a main component of phase noise in the RTDs. The computational times for these studies are sometimes excessive. But the physics indicates that when the device is undergoing self-excited oscillations, it can be characterized and treated as a simple non-linear NDC element, with temporal scaling determined by simple SPICE type algorithms. The Wigner simulation is needed to determine the upper frequency of sustained oscillations and to enhance the understanding of device operation. The Wigner function is also needed to determine the phase noise, because the recovery time depends on the detail time transients of the carriers.

Acknowledgments

This study was supported by the Office of Naval Research.

References

- Grubin H.L. and Buggeln R.C. RTD relaxation oscillations and the time dependent Wigner equation, to be published.
Vergheze S., Parker C.D., and Brown E.R. 1998. Phase noise of a resonant-tunneling relaxation oscillator. *Applied Physics Letters* 72(20): 2550-2552.
Wigner E. 1932. On the quantum correction for thermodynamic equilibrium. *Physical Review* 40: 749-759.



Modeling of Shallow Quantum Point Contacts Defined on AlGaAs/GaAs Heterostructures: The Effect of Surface States

G. FIORI, G. IANNACCONE AND M. MACUCCI

Dipartimento di Ingegneria dell'Informazione, Università degli studi di Pisa, Via Diotisalvi 2, 56122, Pisa, Italy

Abstract. We have developed a program for the simulation of devices defined by electrostatic confinement on the two-dimensional electron gas in AlGaAs/GaAs heterostructures. Our code is based on the self-consistent solution of the Poisson-Schrödinger equation in three dimensions, and can take into account the effects of surface states at the semiconductor-air interface and of discrete impurities in the doped layer. We show results from the simulation of quantum point contacts with different lithographic gaps, whose conductance is computed by means of a code based on the recursive Green's functions formalism.

Keywords: heterostructures, mesoscopic devices, surface states

1. Introduction

The confining potential and the charge density in mesoscopic devices defined by electrostatic confinement in a shallow two-dimensional electron gas (2DEG) strongly depend on the properties of the surface, i.e., on the density of states and the semiconductor-air interface. For this reason, the accurate simulation of such devices requires that proper boundary conditions be enforced at the exposed semiconductor surface (Chen and Porod 1993, Davies and Larkin).

As shown in Iannaccone *et al.* (2000), the assumption of Fermi level pinning at the exposed surface, as well as the assumption of a constant electric field at the semiconductor-air interface, corresponding to a frozen surface charge, are not adequate to achieve results in quantitative agreement with experiments. In particular, for the case of quantum point contacts defined by split gates on an AlGaAs/GaAs heterostructure, these assumptions provide reasonably good results for small lithographic gaps, while for larger gaps do not even reproduce pinch-off of the channel, which is experimentally observed (Iannaccone *et al.* 2000).

A more detailed model of surface states must therefore be used: in particular, we use a model typical of metal-semiconductor contacts (Sze 1981), and based on two parameters: an “effective” work function Φ^* of

the exposed surface, and a constant density of surface states per unit energy per unit area D_s . If E_0 is the energy of the vacuum level, we assume that surface states with energy lower than $E_0 - q\Phi^*$ behave as acceptor states, while surface states with energy higher than $E_0 - q\Phi^*$ behave as donor states.

2. Simulations

We have considered several quantum point contacts defined by split gates on an AlGaAs/GaAs heterostructure, with different lithographic gaps. The layer structure consists of an undoped GaAs substrate, an undoped 12 nm Al_{0.2}Ga_{0.8}As spacer layer, a 31 nm layer of doped GaAs (approx. 10¹⁸ cm⁻³) and an undoped 9 nm GaAs cap layer.

We have solved self-consistently the Schrödinger and Poisson equations in a three dimensional domain in order to obtain the profiles of the first subband and of the electron density in the 2DEG. The potential profile in the three-dimensional structure obeys the Poisson equation

$$\nabla[\varepsilon(\vec{r})\nabla\phi(\vec{r})] = -q[p(\vec{r}) - n(\vec{r}) + N_D^+(\vec{r}) - N_A^-(\vec{r})], \quad (1)$$

where ϕ is the electrostatic potential, ε is the dielectric constant, p and n are the hole and electron densities, respectively, N_D^+ is the concentration of ionized donors and N_A^- is the concentration of ionized acceptors. While hole, acceptor and donor densities are computed in the whole domain with the semiclassical approximation, the electron concentration in the 2DEG is computed by solving the Schrödinger equation with density functional theory.

The observation that electron confinement is strong along the direction perpendicular to the AlGaAs/GaAs interface has led us to decouple the Schrödinger equation into a 1D equation in the vertical (x) direction and a 2D equation in the y - z plane: the density of states in the horizontal plane is well approximated by the semiclassical expression, since there is no in-plane confinement, while discretized states appear in the vertical direction. The single particle Schrödinger equation in 3D reads

$$\begin{aligned} -\frac{\hbar^2}{2} \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} \Psi - \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} \Psi \\ -\frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \Psi + V \Psi = E \Psi; \end{aligned} \quad (2)$$

we can write $\Psi(x, y, z)$ as $\Psi(x, y, z) = \psi(x, y, z) \chi(y, z)$. By substituting the above expression in (2) we obtain the following expression

$$\begin{aligned} -\frac{\hbar^2}{2} \chi \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} \psi - \left[\frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} \right. \\ \left. + \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \psi \chi + V \psi \chi = E \psi \chi, \end{aligned} \quad (3)$$

where the dependence on x , y and z is omitted for clarity. If ψ satisfies the Schrödinger equation along the x direction

$$-\frac{\hbar^2}{2} \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} \psi + V \psi = \tilde{E}(y, z) \psi, \quad (4)$$

by substituting (4) in (3) we obtain

$$\begin{aligned} -\left[\frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} + \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \psi \chi \\ = E \psi \chi - \tilde{E}(y, z) \psi \chi. \end{aligned} \quad (5)$$

Assuming that $\psi(x, y, z)$ is weakly dependent on y and z , and defining

$$\hat{T}_{yz} \equiv -\frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} - \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z}, \quad (6)$$

(5) can be approximated as

$$\psi \hat{T}_{yz} \chi = \psi [E - \tilde{E}_i(y, z)] \chi, \quad (7)$$

where \tilde{E}_i is the i -th eigenvalue of (4). Since $\tilde{E}_i(y, z)$ in the cases considered is rather smooth in y and z , we will assume that eigenvalues of (7) essentially obey the 2D semiclassical density of states.

The confining potential V can be written as $V = E_C + V_{exc}$, where E_C is the conduction band and V_{exc} is the exchange-correlation potential within the local density approximation (Inkson 1984).

$$V_{exc} = -\frac{q^2}{4\pi^2 \varepsilon_0 \varepsilon_r} [3\pi^3 n(\vec{r})]^{\frac{1}{3}} \quad (8)$$

For GaAs, we have $m_x = m_y = m_z = m = 0.067m_0$, where m_0 is the electron mass, therefore the electron density can be written as

$$\begin{aligned} n(x, y, z) = & \frac{k_B T m}{\pi \hbar^2} \sum_{i=0}^{+\infty} |\psi_i(x, y, z)|^2 \\ & \times \ln \left[1 + \exp \left(-\frac{\tilde{E}_i(y, z) - E_F}{k_B T} \right) \right] \end{aligned} \quad (9)$$

where ψ_i and \tilde{E}_i are the eigenfunctions and eigenvalues of (4), respectively.

To solve self-consistently the Poisson-Schrödinger equation, we have used the Newton-Raphson method with a predictor/corrector algorithm close to that proposed in Trellakis *et al.* (1997). In particular, the Schrödinger equation is not solved at each Newton-Raphson iteration step. Indeed, if we consider the eigenfunction constant within a loop and eigenvalues shifted by a quantity $q(\phi - \tilde{\phi})$, where $\tilde{\phi}$ is the potential used in the previous solution of the Schrödinger equation and ϕ is the potential at the current iteration, then the electron density becomes

$$\begin{aligned} n(x, y, z) = & \frac{k_B T m}{\pi \hbar^2} \sum_i |\psi_i(x, y, z)|^2 \\ & \times \ln \left[1 + \exp \left(-\frac{\tilde{E}_i(y, z) - E_F + q(\phi - \tilde{\phi})}{k_B T} \right) \right] \end{aligned} \quad (10)$$

The algorithm is then repeated cyclically until the norm of $\phi - \tilde{\phi}$ is smaller than a predetermined value.

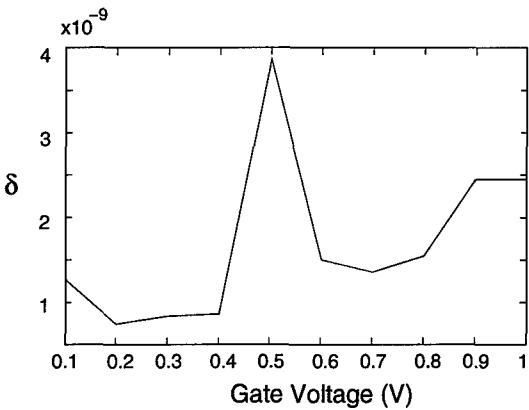


Figure 1. Plot of the parameter δ as a function of the gate voltage V_G for a quantum point contact with lithographic gap of 112 nm.

Once the subband profile is obtained, the conductance in the channel is computed with a method based on recursive Green's functions (Macucci, Galick and Ravaoli 1995).

2.1. Decoupling of the Schrödinger Equation

In order to assess the validity of the approximation which led us to decouple the Schrödinger equation, we define

$$a(x, y, z) \equiv \hat{T}_{yz} \psi_i \chi - \psi_i \hat{T}_{yz} \chi; \quad (11)$$

$a(x, y, z)$ is the difference between the left-hand sides of (5) and (7), and, if the approximation is valid, must be much smaller than the right-hand side in any point of the domain. This means that the term δ , defined as

$$\delta \equiv \max_{x,y,z} \left| \frac{a(x, y, z)}{[E - \tilde{E}_i(y, z)] \phi \chi} \right|, \quad (12)$$

must be much smaller than 1.

In Fig. 1 we plot δ as a function of the voltage applied on the split gates for a quantum point contact with lithographic gap of 112 nm. As can be seen, δ is smaller than 10^{-8} and therefore the approximation is very good.

3. Results

To reach convergence at the desired temperature of 4.2 K, a preventive "cooling" procedure is required,

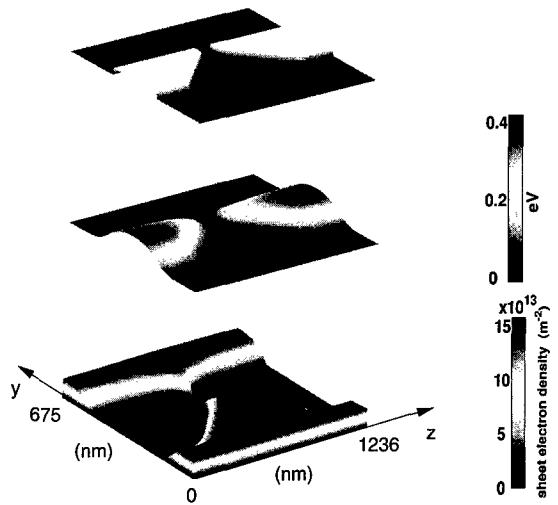


Figure 2. Gate layout of a quantum point contact with lithographic gap of 112 nm (top), theoretical first subband profile (center) and electron density in the 2DEG (bottom).

starting from 100 K, and progressively decreasing the temperature.

The parameters of the surface state model and the concentration of donors in the doped layer have been extracted from measurements on purposely fabricated test structures (Pala *et al.* submitted): $\Phi^* = 4.85$ eV, $D_S = 5 \times 10^{12} \text{ cm}^{-2} \text{ eV}^{-1}$. N_D has been chosen as a fitting parameter in order to reproduce the experimental pinch-off voltages of QPCs with different lithographic gaps. The best fit is provided by $N_D = 0.8 \times 10^{18} \text{ cm}^{-3}$. The electron concentration in the 2DEG is $4 \times 10^{11} \text{ cm}^{-2}$.

In Fig. 2 we plot the gate layout (above), the first subband in the 2DEG (center), and the electron density in the 2DEG (below) for a quantum point contact with lithographic gap of 112 nm and applied voltage of -0.5 V.

Theoretical G - V curves of QPCs with lithographic gap of 57, 112 and 140 nm are shown in Fig. 3. With just one fitting parameter (N_D), computed pinch-off voltages agree within 5% with the average experimental pinch-off voltages measured on the same structures (Fiori *et al.* submitted).

The concentration of impurities in the doped layer plays an important role in the electrical properties of devices realized on a 2DEG (Thean, Nagaraja and Leburton 1997). A simulation that takes into account the random distribution of impurities in the bulk is therefore necessary. In particular we assume that implanted impurities in the bulk obey a Poisson

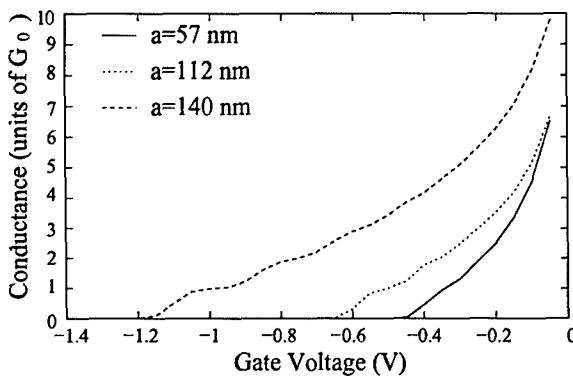


Figure 3. Simulated conductance as a function of gate voltage for devices with lithographic gaps of 57, 112, and 140 nm.

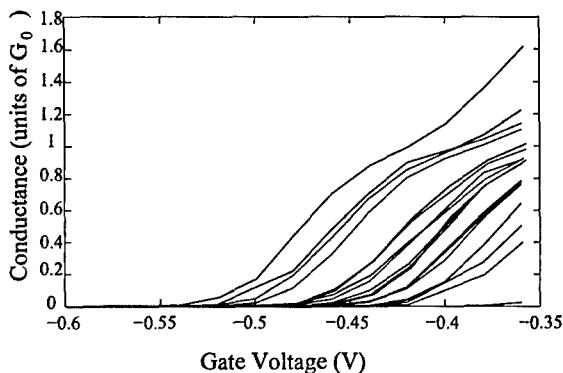


Figure 4. Simulated conductance as a function of gate voltage for 16 nominally identical quantum point contacts with $a = 57\text{ nm}$, but different actual discrete dopant density.

distribution. We have then simulated an ensemble of devices with identical nominal doping profile but different actual distribution of discrete impurities.

Simulated G - V curves of nominally identical quantum point contacts with different ‘actual’ dopant distribution are shown in Fig. 4. For each point of the grid we have considered its associated element of volume ΔV and the nominal doping concentration N_D . The actual number of impurities in ΔV is obtained as a random number N' extracted with Poisson distribution of average $\Delta V N_D$. Dividing N' by ΔV we obtain the actual local density of dopants.

We have obtained a standard deviation of the pinch-off voltage $\sigma_{N_D} = 41.5\text{ mV}$, which is about a half of the experimental value (Fiori *et al.* submitted). Such difference may be due to other sources of dispersion of the pinch-off voltage, such as geometric tolerances.

4. Conclusion

A solver of the Poisson-Schrödinger equations in three dimensions has been developed, which includes a model for surface states based on two parameters: an ‘effective’ work function of the surface states and the density of surface states per unit area per unit energy. We have demonstrated that in the simulation of shallow QPCs the Schrödinger equation may be solved only in the vertical direction, with practically no loss of accuracy.

We have shown that our code can also include the effect of discrete impurities in the doped layer, and that such an effect accounts for about a half of the dispersion of pinch-off voltage measured in experiments.

Acknowledgment

Support from the NANOTCAD Project (IST-1999-10828 NANOTCAD) is gratefully acknowledged.

References

- Chen M. and Porod W. 1993. J. Appl. Phys. 75: 2545.
- Davies J.H. and Larkin I.A. 1994. Phys. Rev. B 49, 4800.
- Fiori G., Iannaccone G., Macucci M., Reitzenstein S., Kaiser S., Kesselring M., Worschech L., and Forchel A. 2002. Nanotechnology 13: 299.
- Iannaccone G., Macucci M., Amirante E., Jin Y., Lanois H., and Vieu C. 2000. Superlattices and Microstructures 27: 359.
- Inkson J.C. 1984. Many Body Theory of Solids—An Introduction. Plenum, New York.
- Macucci M., Galick A., and Ravaioli U. 1995. Phys. Rev. B 52: 5210.
- Pala M., Iannaccone G., Kaiser S., Schliemann A., Worschech L., and Forchel A. 2002. Nanotechnology 13: 373.
- Sze S. 1981. Physics of Semiconductor Devices, 2nd edn. Wiley and Sons, New York.
- Thean V.Y., Nagaraja S., and Leburton J.P. 1997. J. Appl. Phys. 82: 1678.
- Trellakis A., Galick A.T., Pacelli A., and Ravaioli U. 1997. J. Appl. Phys. 81: 7800.



Study of Noise Properties in Nanoscale Electronic Devices Using Quantum Trajectories

XAVIER ORIOLS*, FERRAN MARTÍN AND JORDI SUÑÉ

Departament d'Enginyeria Electrònica—ETSE, Universitat Autònoma de Barcelona, 08193-Bellaterra,
Barcelona, Spain
Xavier.Oriols@uab.es

Abstract. Noise properties in nanoscale devices are studied extending, via quantum trajectories, the classical particle Monte Carlo techniques to devices in which quantum non-local effects are important. This approach can be used to study noise in a wide range of frequencies and can also be easily coupled to a Poisson solver to study long range Coulomb effects in noise characteristics. As a numerical example, we have studied noise in a tunneling barrier showing that the results obtained within our approach exactly reproduce those of the standard Landauer-Buttiker formalism in the zero frequency limit.

Keywords: noise, mesoscopic transport, Bohm trajectories, Monte Carlo technique

1. Introduction

The recent forecast predicts a new generation of electronic devices in the nanometer scale such as 10 nm channel length transistors (Naveh and Likharev 2000). The electrical characteristics of these devices are determined by an interesting interplay between quantum mechanical (QM) and classical theories. Among other nanoscale topics, the noise due to the discreteness of the electron charge has become a very active field of research in mesoscopic devices where classical and quantum knowledge merge together. The Landauer-Buttiker scattering approach has become the standard to study nanoscale devices when phase-coherence is preserved. It provides a transparent description of electron transport, both, for the average current values, \bar{I} , due to Landauer (1957):

$$\bar{I} = \frac{q}{\pi\hbar} \int_0^\infty T(f_L - f_R) dE \quad (1)$$

and for the spectral power of current fluctuations, S_I , mainly due to Buttiker (1990) who, using a second quantization formalism, showed that the spectral power

density for one-dimensional systems at low frequencies can be expressed as:

$$S_I(0) = \frac{q^2}{\pi\hbar} \int_0^\infty \{T\{f_L \cdot (1 - f_R) + f_R \cdot (1 - f_L)\} - T^2 \cdot (f_L - f_R)^2\} dE \quad (2)$$

where q is the absolute value of the electron charge, T is the transmission coefficient as a function of the total electron energy E and $f_{L/R}$ are the Fermi-Dirac occupation functions at the left (right) reservoir related with the chemical potentials at the right and left, $\mu_{L/R}$ (see Fig. 1). (See alternative demonstrations of Lesovik (1989) and Yurke and Kochanski (1990).)

On the other hand, when phase-coherence does not play an essential role, a classical particle description, based on Monte Carlo (MC) techniques, has been used by several authors (González *et al.* 1998, 1999, Korotkov and Likharev 2000) to study fluctuations in mesoscopic systems.

In this letter we present an approach, based on quantum trajectories associated to time dependent wave packets, to study not only the average current, but also current fluctuations in nanoscale devices. It extends the classical MC technique to devices where QM

*To whom correspondence should be addressed.

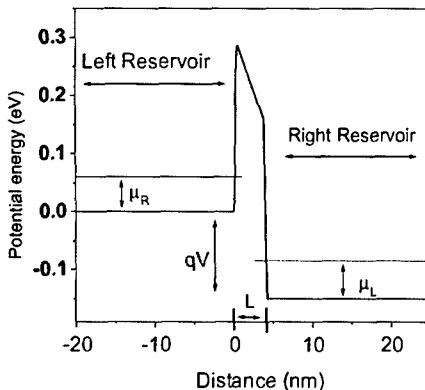


Figure 1. Schematic potential profile considered for the numerical simulation. Electrons are described by Bohm trajectories along the whole simulating box that includes the sample and the two reservoirs.

phase-coherent effects (such as tunneling through a potential barrier) are of prime importance. In particular, the use of Bohm trajectories ensures that the average results of the standard QM theory are exactly reproduced and, at the same time, that the discrete nature of electrons is implicitly considered (Bohm 1952). With our method, the classical MC techniques used to compute current and spectral power density can also be applied in phase-coherent devices. Our approach is useful to study noise in a wide range of frequencies and can also be easily coupled to a Poisson solver to study the effects of long range Coulomb interaction between carriers in the noise characteristics of tunnel devices. Our work is based on previous ideas of Landauer (Martin and Landauer 1992, Landauer 1989) who studied shot noise within a wave-packet point of view.

2. Noise in Terms of Wave Packets

Although the Buttiker formalism (1990) has become the standard to study noise in coherent devices, other *more-intuitive* approaches have also successfully explained noise characteristics in these devices. Among others, together with Th. Martin (and with the goal of reproducing the Buttiker results), Landauer provided a simple derivation of the spectral power density of the shot noise (i.e. expression 2) within a time dependent wave packet picture (Landauer 1989). In the following, part of his analysis will be repeated for convenience. Let us assume a one-dimensional system with quantum ballistic transport. For each small energy interval, ΔE , the current can be represented as a set of δ -pulses of area $\pm q$ which account for the *random* transmission

of electrons between the reservoirs. A positive pulse is measured when a wave packet incident from the left electrode is transmitted to an empty state in the right reservoir. The probability of this event is $T \cdot f_L(1 - f_R)$, where $1 - f_R$ factor accounts for the Pauli principle. The transmission from right to left gives a negative pulse whose probability is $T \cdot f_R(1 - f_L)$. For a single injection event, the average value of charge detected at the right contact is defined as:

$$\bar{Q} = (q) \cdot T \cdot f_L(1 - f_R) + (-q) \cdot T \cdot f_R(1 - f_L) \\ = q \cdot T \cdot (f_L - f_R)$$

In order to compute the noise as the standard deviation of Q , we compute the square average value of the measured charge:

$$\overline{Q^2} = q^2 \cdot (Tf_L(1 - f_R)) + (-q)^2 \cdot (Tf_R(1 - f_L))$$

In this regard, the power of the current fluctuations can be computed as:

$$\overline{\Delta Q^2} = \overline{Q^2} - (\bar{Q})^2 = q^2 \cdot (Tf_L(1 - f_R)) \\ + (-q)^2 \cdot (Tf_R(1 - f_L)) - (qT(f_L - f_R))^2$$

To obtain $S(0)$ we just have to integrate $\overline{\Delta Q^2}$ over the whole energy range multiplied by the one-dimensional density of incoming electrons (which in our one dimensional case can be computed as $v = dE/\hbar\pi$ (Irmy 1997)). By doing this integration, expression 2 is exactly reproduced. As we will show in this work, this alternative picture for electronic noise is quite *naturally* supported within Bohm interpretation of QM (Bohm 1952).

3. Our Model

During last years, the research of our group has been focused on extending the classical MC techniques to quantum devices where phase coherence plays an important role. Focused on the resonant tunneling devices (which have a rich QM phenomenology), we have developed a quantum MC formalism and we have obtained self-consistent results for the average current (Oriols *et al.* 1998, 1999). Our present approach is an extension of that previous work where we deal, not only with average values, but with fluctuations. In this regard, in order to be able to compute noise characteristics from quantum MC simulations special attention has to be devoted to two points: the injection

statistics and the measurement of the current. After a brief introduction to the use of Bohm trajectories for the simulation of electronic transport in mesoscopic devices, in this section we will focus on these two topics.

Bohm's interpretation of QM exactly reproduces the statistical predictions of the standard Copenhagen interpretation and, at the same time, provides a causal description for the individual behavior of QM systems. Within the Bohm's interpretation, all the particles of a quantum pure state ensemble follow different and well-defined causal trajectories under the combined influence of the classical potential, $V(x, t)$, and a new term called the quantum potential, $Q(x, t)$, which is directly related to the wave function (Bohm 1952). In order to compute Bohm trajectories, first, the time-evolution of a wave packet $\Psi(x, t)$, solution of the time-dependent Schrodinger equation, must be known. Then, according to Bohm approach, the instantaneous velocity, $v(x, t)$, for an electron located at position x and time t is given by $v(x, t) = J(x, t)/|\Psi(x, t)|^2$ where $J(x, t)$ is the quantum mechanical particle current density. The electron causal trajectory, $x = x(x_0, t)$, is determined by integrating $v(x, t)$ after fixing its initial position x_0 . This initial position accounts for the unavoidable uncertainty in QM and is randomly selected according to the probability $|\Psi(x_0, 0)|^2$. Let us notice that the main difference between a classical MC scheme and our proposal lies in the expression used to compute electron velocity: in the former, the velocity is proportional to the local electric field, while in our approach, the electron velocity takes into account the QM non-local effects via $\Psi(x, t)$. The detailed procedure that we use for the computation of Bohm trajectories has been published elsewhere (Oriols *et al.* 1998, 1999).

Let us move to the injection model of our QM simulator. When dealing with mesoscopic device simulations, the modeling of carrier injection from thermal reservoirs is a delicate problem. According to Levitov and Lesovik (1993) and Levitov, Lee and Lesovik (1996), under degenerate conditions one should use a Binomial distribution instead of a Poissonian one. An injection model for MC particles has been developed by González *et al.* (1999) showing its accuracy to describe either non-degenerate or completely degenerate conditions in one-dimensional mesoscopic conductors. In our quantum MC simulator, we will use that model. As we have previously noticed, the rate of incoming electrons impinging with velocity v_0 upon the boundary between the leads and the conductor, v , is given by

the product of their velocity and the one dimensional density of states $v = v_0 dk/\pi = h k_c dk/m^* \pi = dE/h\pi$ (where we have taken a parabolic isotropic relation for the energy-momentum relationship and an effective mass equal to m^*). On the other hand, the injection model has to take into account the occupation function at the leads. In this regard, the probability of injecting a wave packet with a positive central momentum k_c depends on the probability of occupation in the left reservoir, f_L , and also on the probability that there is no wave packet with the same central momentum k_c at the right contact, $1 - f_R$. This point differs from González's model that deals with point particles and only considers the occupation function in the left reservoir. Our algorithm to inject particles from the left contact with velocity v_k is the following: At each time interval of duration, $v^{-1} = m^* \pi / h k_c \Delta k$ an attempt to introduce a wave packet takes place (in our case, Δk is the inverse if the wave packet spatial dispersion $\Delta k = 1/\sigma$). Then, a random number r uniformly distributed between zero and one is generated, and the attempt is considered successful only if $r < f_L \cdot (1 - f_R)$. Similar arguments are used to inject electrons from the right reservoir. This procedure exactly takes into account the injection noise of the system (González *et al.* 1999). Then, each time that an electron is definitively injected in the simulating box, its initial position is selected according to the probability presence $|\Psi(x_0, 0)|^2$ (Oriols *et al.* 1998). The uncertainty in the initial position is transferred to an uncertainty in the transmittance (i.e. there are electrons that can pass through the barrier and others that are reflected). This additional random selection takes into account the partition noise due to the barrier.

The second topic that we want to address is the measurement of the current. The meaning of measurement in QM carries some difficulties related to the behavior of the wave function during the measurement process. However, in our approach, since we deal with causal trajectories, current can be computed following classical MC techniques. In particular, according to the extension of Ramo-Shockley theorem to semiconductor devices (Cavalleri *et al.* 1971, Pellegrini 1986), the total instantaneous current $I(t)$ through each cross sectional area of the device, the sum of conduction and displacement current, is computed as:

$$I(t) = \frac{q}{L} \sum_{i=1}^{N(t)} v_i(x, t) \quad (3)$$

where L is the length of the device, $N(t)$ is the total number of carriers which are instantaneously inside the device, and $v_i(x, t)$ is the value of the Bohm velocity at time t and position x . The level i identifies each electron, and only those within $0 < x < L$ are considered (see Fig. 1). Once the current is recorded for a sufficient long period of time (in our simulation 50 ps), the power spectral density, $S(w)$, of the current fluctuations can be computed by Fourier transforming the autocorrelation function of the current fluctuations following standard classical MC methods (Varani *et al.* 1994).

4. Numerical Results

In order to show the capabilities of this approach we will provide a numerical example of noise in single tunneling barrier devices. We will focus on low frequency noise to compare our results with the standard Buttiker formalism. We will consider ballistic electronic transport in a one-dimensional tunneling barrier. Our example, schematically described in Fig. 1, consists in two highly doped layers of AsGa separated by a layer of $\text{Al}_x\text{AsGa}_{1-x}$ that introduces a 0.3 eV potential barrier height. We assume that the applied bias, V , falls only in the barrier region without voltage fluctuations in the contacts. The two AsGa layers are considered large enough to be characterized as perfect reservoirs with the Fermi-Dirac distribution $f_{L/R}$ at 300 K with $\mu_L = \mu_R + q \cdot V$. We consider injection from both reservoirs, left and right, but from a unique energy. In this regard, we define two wave packets with the same central energy, $E = 0.15$ eV, but different initial central positions and opposite central wave vectors. At time $t = 0$, the initial probability presence of each wave packets, $|\Psi(x, 0)|^2$ corresponds to a Gaussian wave packet with a spatial dispersion $\sigma_x = 130$ Å, which is much longer than the sample length ($L = 40$ Å). The wave packet evolution is calculated by solving the Schrödinger equation along a simulating box of 2048 Å, that includes the sample and the two reservoirs (a unique effective mass equal to 0.067 times the free electron mass is used). In order to compare our numerical results with Buttiker formalism (i.e. with Eq. 2), the potential profile is considered to be time independent without Poisson self-consistency (see Fig. 1).

In Fig. 2, the simulated values of the current obtained from a total simulation time of 50 ps and $\Delta T = 0.25$ fs, are represented. The instantaneous current is

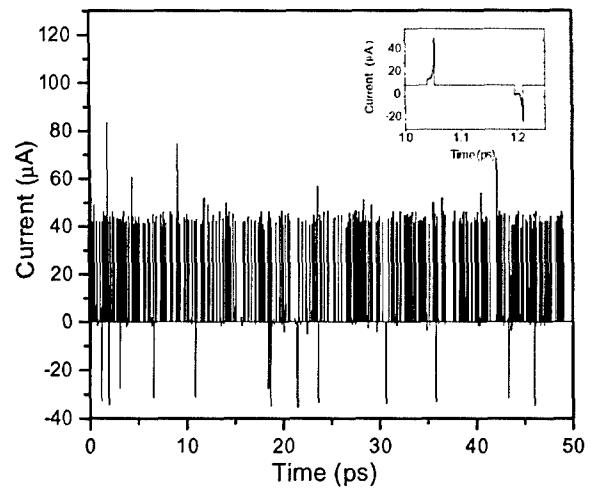


Figure 2. (a) Instantaneous current for the device described in Fig. 1 with an applied bias $V = 0.075$ Volts and $\mu_R = \mu_L = 0.2$ eV. The inset shows two electron pulses.

computed from Eq. 3 and shows positive and negative one-electron pulses associated to left and right injection (see inset of Fig. 2). Each pulse corresponds to an electron that spends a time L/v_0 to traverse the barrier. The velocity of electrons is not exactly constant mainly because of the applied potential that provides electrons with a higher velocity near the right contact. Following the standard procedure (Varani *et al.* 1994), the one-side spectral noise power of the current fluctuations, $S(w)$, can be computed from $I(t)$. In Fig. 3 we see that $S(w)$ has a constant value for low frequencies

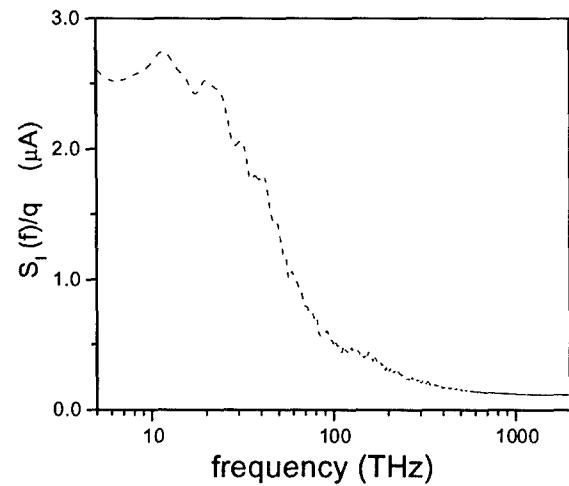


Figure 3. The one-side noise spectral power density $S_I(w)$ for the sample of Fig. 1 for $V = 0.15$ Volts and $\mu_R = \mu_L = 0.2$ eV.

and starts to decrease at a frequency associated to the electron transit time across the sample (in our case the transit time is 20 fs and the cut-off frequency 50 THz). In order to test the validity of our approach, we compare our results with those obtained from the Buttiker formalism. The comparison is carried out in terms of the Fano factor, F , defined by $S_1(0) = F \cdot 2 \cdot q \cdot I$. The analytical results are computed from Eqs. (1) and (2) considering a mono-energetic system where T is defined as the average transmission coefficient of the wave packets (Leavens and Aers 1993). The numerical results are computed by repeating the results of Fig. 3 for different applied bias. In Fig. 4(a) we have plotted the transmission coefficient and the left/right occupation functions f_L/f_R for the different applied bias. Since the applied bias lowers the effective barrier, T grows with V . On the other hand, for high voltages f_R is so low that only injection from the left reservoir is representative. In Fig. 4(b), we have compared the Fano factor between our approach (squares) and

Buttiker formalism (circles). The excellent coincidence for the Fano factor shows the viability of using Bohm trajectories in a MC scheme for studying noise characteristics in phase-coherent mesoscopic devices. For low bias, the transmittance through the barrier is so low, that the electrons follow a Poisson distribution. On the other hand, as we have said, for high voltages only the injection from the left reservoir must be considered, but since the transmission coefficient of the barrier is moderately high, shot noise following a binomial partition process appears. Hence, the expected Fano factor approximates $F = 1 - f_L \cdot T$. It is interesting to notice that, as we see in Fig. 4, for f_R close to f_L the Fano factor can be greater than 1 (this situation means a very low current). The equivalence between both approaches for the low frequency limit is not surprising since, as we have seen in the introduction, $S(0)$ can be deduced just with probability argument for the partition noise of the barrier and the injection noise (Landauer 1989, Irmy 1997).

5. Conclusions

In conclusion, we have developed a MC simulator for phase-coherent mesoscopic devices by means of Bohm trajectories associated to time dependent wave packets to describe the electron path. Our approach is based on two fundamental characteristics of the Bohm's approach: the average QM results (such as average current or transmission coefficient) are perfectly reproduced in terms of Bohm trajectories; and the discrete nature of electrons is explicitly considered in Bohm's formulation (allowing noise computation using classical techniques). In this regard, this work follows the path opened by Martin and Landauer (1992), Landauer (1989) who deduced Buttiker formalism within a simple wave packet framework. The main potentialities of our approach are related with its capability to include a Poisson solver to obtain self-consistent potential profiles and noise spectra at high frequencies. These conditions are not easily accounted for in present phase-coherent noise theories and drastically modify noise characteristics.

Acknowledgment

The authors are really grateful to Javier Mateos, Tomás González and Daniel Pardo for helpful discussion. This work has been partially supported by the Dirección General de Enseñanza Superior e Investigación

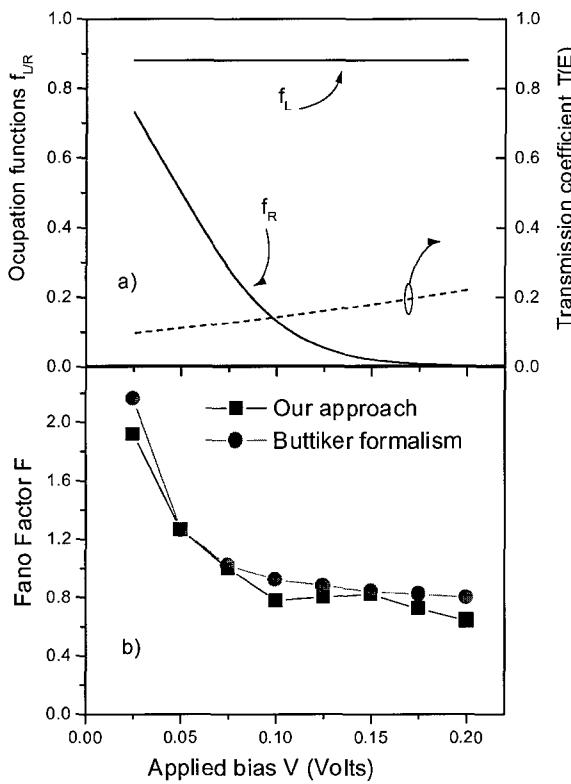


Figure 4. Noise characteristics as a function of the applied bias V : (a) The transmission coefficient T and occupation functions $f_{L/R}$ (b) the Fano factor F computed within our model (squares) and within Buttiker formalism (circles).

through project BFM2000-0353 and by a grant of the *Programa científico de la OTAN*.

References

- Bohm D. 1952. Phys. Rev. 85: 166.
- Büttiker M. 1990. Phys. Rev. Lett. 65: 2901.
- Cavalleri G., Gatti E., Fabbri G., and Svelto V. 1971. Nucl. Instrum. Methods 92: 137.
- González T., González C., Mateos J., Pardo D., Reggiani L., Bulashenko O.M., and Rubí J.M. 1998. Physical Review Letters 80: 13.
- González T., Mateos J., Pardo D., Varani L., and Reggiani L. 1999. Semicond. Sci. Technol. 14: L37–L40.
- Irmy Y. 1997. Introduction to Mesoscopic Physics. Oxford University Press, New York, p. 98.
- Korotkov A.N. and Likharev K.K. 2000. Phys. Rev. B 61(23): 15975.
- Landauer R. 1957. IBM J. Res. Dev. 1: 223.
- Landauer R. 1989. Physica D. 38: 226.
- Leavens C.R. and Aers G.C. 1993. In: Wiesendanger R. and Güntherodt H.-J., Scanning Tunneling Microscopy III. Springer, New York.
- Lesovik G.B. 1989. JETP Lett. 49: 592.
- Levitov L.S., Lee H., and Lesovik G.B. 1996. J. Math. Phys. 37: 4845.
- Levitov L.S. and Lesovik G.B. 1993. JETP Lett. 58: 230.
- Martin Th. and Landauer R. 1992. Phys. Rev. B 45: 1742.
- Naveh Y. and Likharev K.K. 2000. IEEE Electron Device Letters 21: 242.
- Oriols X., Garcia J.J., Martín F., Suñé J., González T., Mateos J., and Pardo D. 1998. Appl. Phys. Lett. 72(7): 806.
- Oriols X., Garcia J.J., Martín F., Suñé J., González T., Mateos J., Pardo D., and Vanbesien O. 1999. Semicond. Sci. Technol. 14: 532.
- Pellegrini B. 1986. Phys. Rev. B 34: 5921.
- Varani L., Reggiani L., Kuhn T., González T., and Pardo D. 1994. IEEE Trans. Electron Devices, 41: 1916.
- Yurke B. and Kochanski G.P. 1990. Phys. Rev. B 41: 8184.



Monte-Carlo Simulation of Clocked and Non-Clocked QCA Architectures

L. BONCI, M. GATTOBIGIO, G. IANNACCONE AND M. MACUCCI

Dipartimento di Ingegneria dell'Informazione, Università degli studi di Pisa, Via Diotisalvi 2, I-56126 Pisa, Italy

Abstract. We present a Monte Carlo simulation of two implementations of Quantum Cellular Automaton (QCA) circuits: one based on simple ground state relaxation and the other on the clocked cell scheme that has recently been proposed by Tóth and Lent. We focus on the time-dependent behavior of two basic circuits, a binary wire and a majority voting gate, and assess their maximum operating speed and temperature requirements for different sets of fabrication parameters.

Keywords: QCA circuits, nanoelectronics ,Coulomb Blockade

1. Introduction

Quantum Cellular Automata (QCA) represent an original approach, first proposed by Lent *et al.* (1993), to the implementation of logic circuits, exploiting the bistable properties of a cell made up of 4 quantum dots or nodes and containing 2 excess electrons. The initial proposal of QCA circuits was based on two-dimensional arrays of such cells and on letting the system relax down to the ground state, so that the result of the computation was obtained as the state of a group of cells located along the boundary of the array. However, if ground state computation is performed with relatively large QCA arrays, the evolution of the system can get temporarily stuck in a metastable state and reach the ground state (and thus the correct logical output) only after an extremely long time (Landauer 1994). To avoid this problem, an adiabatic logic scheme has been proposed, in which the evolution of the system is driven by a multi-phase clock (Lent and Tougaw 1997). This scheme involves modulation of the interdot barriers, in order to keep each cell always in its instantaneous ground state and to lock it, i.e. freeze its state, before it is used to drive a neighboring cell.

An interesting approach to the modulation of the inter-dot barriers in a metal-dot QCA implementation has been proposed in Tóth and Lent (1999): it consists in implementing the barrier with two additional dots, whose potential can be varied by means of an external voltage.

Metastable states are no longer a problem for adiabatic logic, but we need to consider that signal propagation is limited by the switching time of single cells. In particular, proper operation can be obtained only if tunneling transition rates are large enough to allow electrons to actually tunnel into the expected dot during the active state of the cell. In Bonci, Iannaccone and Macucci (2001) we evaluated the switching time by computing the electron tunneling rates as a function of material parameters and cell geometry. Here we use such results to test the operation of circuits made up of non-clocked and clocked cells via a Monte Carlo simulation and we compare the achievable performance.

2. QCA Circuit Simulator

The numerical simulation has been performed by means of a Monte Carlo code that we have developed, based on the orthodox Coulomb Blockade theory and specifically suited to handle circuits containing clocked single-electron devices (Macucci, Gattobigio and Iannaccone to appear). Our software allows simulating circuits with voltage sources that have an arbitrary piecewise linear time dependence.

In addition, since cotunneling plays an important role in some regimes of operation of QCA circuits, it has been taken into consideration, although approximately, on the basis of the formulation in Fonseca *et al.* (1995).

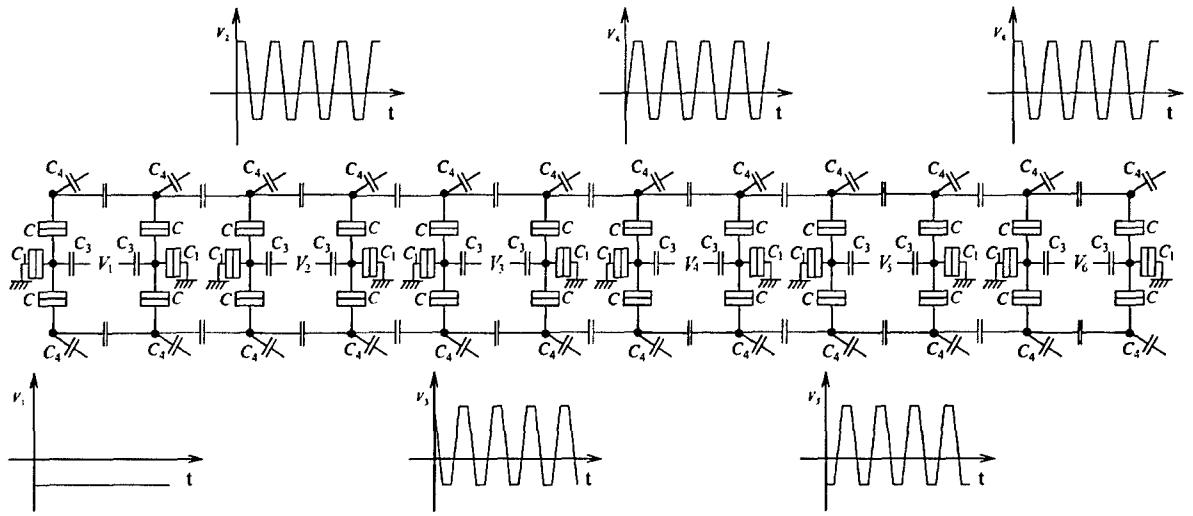


Figure 1. Chain made up of six clocked cells.

We study two different circuits: a linear chain made up of six QCA cells and a majority voting gate made up of eight QCA cells, both of them in the clocked and non-clocked version. The clocked chain, relaying on relaxation down to the ground state, is shown in Fig. 1. The non-clocked versions are much simpler: each cell is made up of a square whose sides are represented by four tunneling capacitors (C_o for the horizontal sides and C_v for the vertical sides), neighboring cells are connected via ideal capacitors C_c and the state of the first cell is enforced via C^* capacitors connected to the two outer nodes of the input cells (Bonci, Iannaccone and Macucci to be published).

In the two cases we use a different simulation strategy, due to the different principle of operation. In the non-clocked case we are strictly following the ground-state calculation paradigm: to obtain the logical output of the circuit we need to wait until the system has relaxed to the ground state. By repeating several times the simulation, we are able to verify circuit reliability (i.e. whether or not the correct logical output is achieved after a given time) and to compute the average time the circuit needs to reach the correct logical output.

In the clocked case the switching time is imposed from the outside. In order to verify circuit operation we follow the time evolution of each single cell and verify whether it is in the expected logical state during the proper time intervals. We do not need to perform an ensemble average in this case, because statistics are obtained over a large enough number of clock cycles.

We have performed calculations for two sets of system parameters. The first set has been derived from the recent experiments which have successfully demonstrated operation of simple QCA gates (Orlov *et al.* 1999, Amlani *et al.* 1999, 2000). The second one was obtained from our previous work, in which we discussed the limits of clocked QCA devices (Bonci, Iannaccone and Macucci 2001) from a theoretical point of view. The experimental and the theoretical choices of parameters are shown in the Tables 1 and 2, where, for the clocked case, C_c represents the coupling capacitor between neighboring cells. The theoretical set represents a compromise between miniaturization, efficiency and technical feasibility, at least in perspective, since fabrication of the corresponding extremely small and precise structures is not yet achievable with current technology.

Table 1. Circuit parameters for the non-clocked case.

	C_o	C_v	C_c	C^*	R_T
Experimental parameters (Orlov <i>et al.</i> 1999, Amlani <i>et al.</i> 1999, 2000, Orlov <i>et al.</i> 2000)	400 aF	288 aF	88 aF	1 aF	200 k Ω
Theoretical parameters (Bonci, Iannaccone and Macucci 2001)	5.3 aF	5.3 aF	2 aF	0.1 aF	200 k Ω

Table 2. Circuit parameters for the clocked case.

	C	C_1	C_2	C_3	C_4	C_c	R_T
Performed experiments (Orlov <i>et al.</i> 1999, Amlani <i>et al.</i> 1999, 2000, Orlov <i>et al.</i> 2000)	420 aF	300 aF	25 aF	80 aF	200 aF	50 aF	200 k Ω
Theoretical prediction (Bonci, Iannaccone and Macucci 2001)	5.3 aF	1.2 aF	1.2 aF	3.57 aF	1.48 aF	1 aF	200 k Ω

3. Binary Wire Simulations

We start by simulating a binary wire based on the experimental parameters. In the non-clocked case the relaxation time to the ground state is a statistical quantity whose average value $\langle t_{rel} \rangle$ is shown in Fig. 2, as a function of temperature. We notice that $\langle t_{rel} \rangle$ decreases as temperature increases. This is due to the increased tunneling rate, which helps driving the evolution of the system out of metastable states, but this phenomenon is limited by the fact that beyond a certain temperature fluctuations prevent the binary wire from reaching a stable ground state at all.

Knowledge of the average quantity $\langle t_{rel} \rangle$ is not sufficient to assess the speed of the circuit. We need to take into consideration the distribution of relaxation times, which is quite broad and exhibits long tails: this implies that $\langle t_{rel} \rangle$ is actually too conservative an estimate. We notice also that, as temperature decreases, the importance of cotunneling events increases.

The situation is similar if we choose the other parameter set, which we have defined as ‘‘theoretical’’. The overall behavior is comparable with the previous parameter choice, with the only significant difference consisting in the possibility to achieve a higher operating temperature.

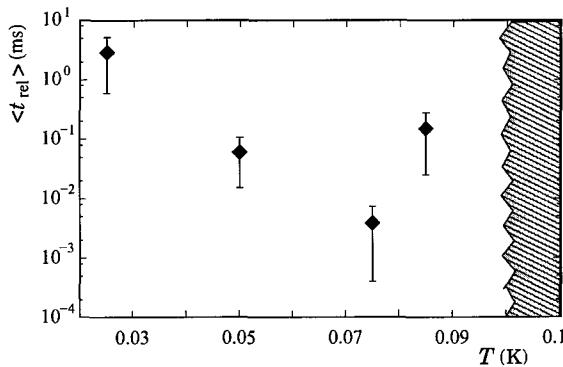


Figure 2. Average relaxation time $\langle t_{rel} \rangle$ as a function of temperature. In the dashed region the operation is completely disrupted due to thermal fluctuations. The values of cell parameters are shown in the first row of Table 1.

Let us now consider the clocked architecture of Fig. 1: by means of an external clock signal, we enforce a well defined switching time, which is not dependent on the relaxation to the ground state any more. In this case we need to assess whether the system is fast enough to follow the clock and thereby to provide the correct final output. The error probability depends both on the clock rate and on the operating temperature. With too fast a clock the cell is not able to switch properly, and the same may happen for too large values of the temperature.

We performed runs over several (250) clock cycles and we checked the logical state of the second and of the sixth (last) cell of the chain. In this way, we were able to compute the percentage of correct output P_{co} over the total number of clock cycles. With this prescription, we obtained the result shown in Fig. 3. The deterioration of the circuit evolution with increasing clock frequency is clear; moreover P_{co} decreases as we move along the chain, due to the fact that the error probability increases as the number of cells that have processed the information increases.

We repeated the simulation with the theoretical parameter set obtaining the results shown in Fig. 4. Similar comments apply to these results, although

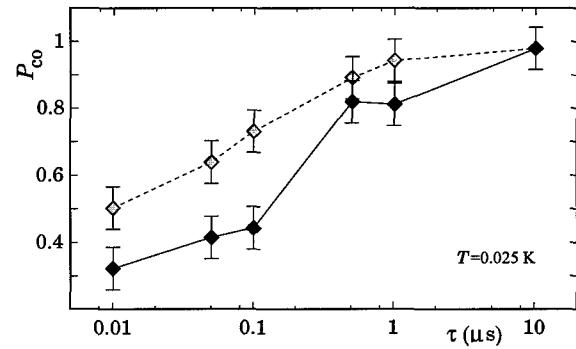


Figure 3. Probability of correct operation (P_{co}) for the second and the last cell in the clocked chain of Fig. 1 as a function of the clock period. The solid curve refers to the last cell while the dashed one refers to the second cell. The values of cell parameters are shown in the first row of Table 2.

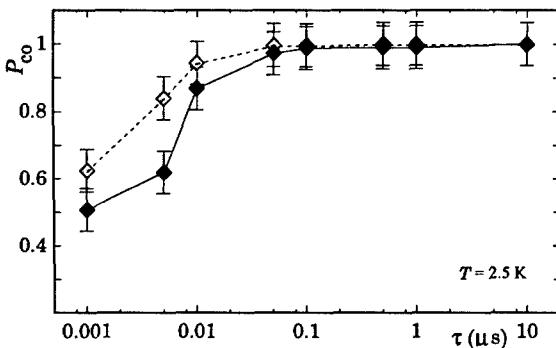


Figure 4. Probability of correct operation (P_{co}) for the second and the last cell in the clocked chain of Fig. 1 as a function of clock period. The solid curve refers to the last cell while the dashed one refers to the second cell. The values of cell parameters are shown in the second row of Table 2.

the temperature range is different (we moved up two orders of magnitude). In this case we can find a region of correct operation extending down to a clock period of 5×10^{-8} s and to a temperature of 2.5 K. This could be an acceptable operating condition, at least for some niche application. This result is worth of discussion. A rough dimensional analysis, based on the $R_T C^{-1}$ time constant, would give a typical switching time of 10^{-12} s, much shorter than the one obtained by means of numerical simulation. There are indeed several phenomena that degrade circuit operation. The actual probability for an electron located in the central dot of a cell to switch to a side dot can be obtained by considering the difference between the voltage drops on the upper and lower tunneling junctions. With our choice of parameters the voltage difference due to another electron located in the other half of the same cell is 3 mV and corresponds to a current of 2.5×10^{11} electrons per second, i.e., to a switching time 4 times larger than the previous estimate. This is true for intracell switching, but we need to consider switching due to the influence of a neighbor cell. The voltage unbalance in this case is typically 5 times smaller and thus the switching time has to be increased by the same factor. A further multiplying factor comes out from the clock time pattern. In order to inhibit unwanted transitions, the locked and null state need to have an energy difference with respect to the active state much larger than $K_B T$, therefore the active region represents only one tenth of the rising segment of the control voltage. Moreover, this segment represents 1/4th of the clock pattern and thus the active region is restricted to a time interval which is 40 times smaller than the clock period. Finally, we need to consider that a clock rate in a logical circuit can be

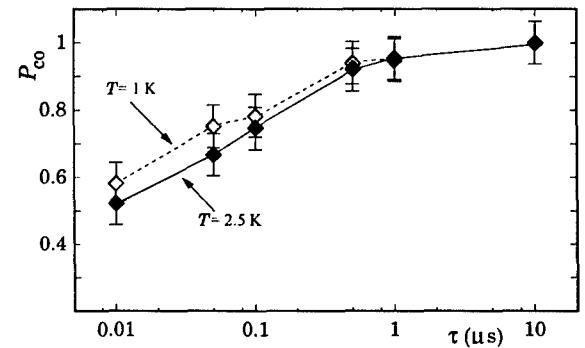


Figure 5. Probability of correct operation (P_{co}) for the output cell in a clocked majority voting gate as a function of clock period. We studied two different temperatures: $T = 1$ K (dashed line) and $T = 2.5$ K (solid line). The values of cell parameters are shown in the second row of Table 2.

considered safe if it is at least ten times smaller than the maximum theoretical rate.

The overall multiplying factor stemming from these considerations approaches 10^4 , bringing us very close to the numerical results.

4. Simulation of a Majority Voting Gate

In the previous section we studied a simple QCA wire, the basic element of QCA logic. A further step consists in considering a circuit that performs a slightly more complex logical operation, such as a majority voting gate. This circuit is expected to provide at the output the logical state which is present at the majority of the inputs. We start with the simulation of a non-clocked circuit, computing $\langle t_{rel} \rangle$ as the average over several realizations. We find results similar to those for the binary wire, with an increase of the average relaxation time, as should be expected as a consequence of the greater circuit complexity.

Finally, we study a clocked majority voting gate and report the results in Fig. 5, for the theoretical choice of parameters. If we compare Fig. 5 with Fig. 4 we notice, as in the case of the non-clocked version, an overall increased error probability, which further limits the maximum operating speed.

5. Conclusions

We have investigated the time-dependent behavior of clocked and non-clocked QCA circuits, obtaining results for the maximum operating speed and operating temperature for two choices of parameters. In the

non-clocked circuits, relaxation to the ground state is a statistical process with a broad distribution characterized by long tails. Even with the theoretical set of parameters, which corresponds to a conceivable, although not yet feasible, implementation of QCA cells, the maximum speed that can be achieved is unsatisfactory, on the one side because of the action of intermediate metastable states into which the evolution of the system gets trapped at low temperature and, on the other side, because of the disrupting action of thermal fluctuations at higher temperatures.

The clocked architecture allows to overcome the problem of metastable states and to achieve much faster operation, with improved control of data flow. We have shown that, with a very optimistic choice of parameters, it is possible to achieve clock frequencies and operating temperatures that can be acceptable for some niche application in which the other advantages of QCA systems may play a role. The layout complexity is, however, very significantly increased moving to the clocked architecture, due mainly to the need for clock distribution lines, which makes practical implementation very challenging.

Acknowledgments

We acknowledge financial support from the European Commission (project **Answers** n. 28667) and from

the Italian National Research Council (project 5% Nanotecnologie).

References

- Amlani I., Orlov A.O., Kummaruku R.K., Bernstein G.H., Lent C.S., and Snider G.L. 2000. *Appl. Phys. Lett.* 77: 738.
- Amlani I., Orlov A.O., Toth G., Bernstein G.H., Lent C.S., and Snider G.L. 1999. *Science* 284: 289.
- Bonci L., Iannaccone G., and Macucci M. 2001. *J. Appl. Phys.* 89: 6435.
- Bonci L., Gattobigio M., Iannaccone G., and Macucci M. Simulation of the time evolution of clocked and nonclocked Quantum Cellular Automaton (QCA) circuits. *J. Appl. Phys.* (to be published).
- Fonseca L.R.C., Korotov A.N., Likharev K.K., and Odinstov A.A. 1995. *J. Appl. Phys.* 78: 3238.
- Landauer R. 1994. In: Welland M.E. (Ed.), *Ultimate Limits of Fabrication and Measurement*. Kluwer, Dordrecht.
- Lent C.S. and Tougaw P.D. 1997. *Proc. IEEE* 85: 541.
- Lent C.S., Tougaw P.D., and Porod W. 1993. *Appl. Phys. Lett.* 62: 714.
- Macucci M., Gattobigio M., and Iannaccone G. 2001. *J. Appl. Phys.* 90: 6428.
- Orlov A.O., Amlani I., Kummamuru R.K., Ramasubramanian R., Toth G., Lent C.S., Bernstein G.H., and Snider G.L. 2000. *Appl. Phys. Lett.* 77: 295.
- Orlov A.O., Amlani I., Toth G., Lent C.S., Bernstein G.H., and Snider G.L. 1999. *Appl. Phys. Lett.* 74: 2875.
- Tóth G. and Lent C.S. 1999. *J. Appl. Phys.* 85: 2977.
- Ungarelli C., Francaviglia S., Macucci M., and Iannaccone G. 2000. *J. Appl. Phys.* 87: 7320.



A Wigner Function Based Ensemble Monte Carlo Approach for Accurate Incorporation of Quantum Effects in Device Simulation*

L. SHIFREN[†] AND D.K. FERRY

*Center for Solid State Electronics Research and Department of Electrical Engineering,
Arizona State University, Tempe, AZ 85287-5706, USA*

lucian.shifren@asu.edu

Abstract. We present results of both Gaussian wave-packet tunneling through a single barrier structure and RTD operation achieved from a particle-based Ensemble Monte Carlo (EMC) simulation that is based on the Wigner distribution function (WDF). Methods of including the Wigner potential into the EMC, to incorporate naturally quantum phenomena, via a particle property we call the affinity are discussed. Results showing tunneling and correlation build-up in both cases are presented.

Keywords: Wigner function, Monte Carlo, resonant tunneling diode

Current device technologies are already at, or quickly approaching, the scales whereby quantum effects due to the strong confinement of carriers and direct source-drain tunneling will begin to dominate (Kawaura *et al.* 1997, Ferry 1985). Ensemble Monte Carlo (EMC) has always been the most vigorous and trusted method for device simulation as it has again and again proven to be reliable as well as predictive. However, EMC relies on the particle nature of the electron, but quantum mechanical phenomena arise from the wave-like nature of the electron. In order to resolve quantum mechanical effects, the wave-like nature of the electron needs to be incorporated into the EMC. To do this, we use the obvious similarities between the Boltzman transport equation and the Wigner function transport equation (Ferry and Grubin 1995, Wigner 1932). While not frequently used, the Wigner distribution function has had success in modeling resonant tunneling diodes (RTD) (Kluksdahl *et al.* 1989, Ravaioli *et al.* 1985, Frenley 1987). To incorporate the WDF, and most importantly, the Wigner potential (which is a non-local potential and is responsible for the quantum effects) into the EMC, we assign an additional property to the electrons that

we term the affinity. The affinity can have any value whose magnitude less than 1, which allows us to easily incorporate any fractional or negative values which the WDF may acquire. By maintaining the essence of the EMC, we allow the particle nature of the EMC to survive and we can then study quantum mechanical effects in the simulation. Other methods have been developed to incorporate the WDF into an EMC (Jacoboni *et al.* 2001, Garcia-Garcia *et al.* 1998). However, our method depends on calculating the Wigner potential exactly and updating the electron (particle) distribution within the standard EMC to account for this non-local correction to the density in the system. Although other, non-WDF methods to include quantum effects have been developed, such as the effective potential (Shifren, Akis and Ferry 2000, Akis *et al.* 2001), these methods account for certain phenomena associated with the wave-like nature of the electron but cannot account for tunneling, correlation or interference effects (Ferry *et al.*).

As mentioned, we assign the particles in the EMC a new property that we call the affinity. The affinity is a value, whose magnitude is less than 1, that the particles carry which represents the particle's contribution to the entire electron distribution. With this, we are able to construct the Wigner function from the particle

*Work supported by the office of Naval Research.

[†]To whom correspondence should be addressed.

distribution using

$$f(x, k) = \sum_i \delta(x - x_i) \delta(k - k_i) A(i), \quad (1)$$

where $A(i)$ is the particle affinity and the delta functions represent the existence of a classical particle. This distribution function can then be used to calculate the non-local Wigner potential (NLP), given as

$$\frac{df(x, k, t)}{dt} = \frac{1}{2\pi\hbar} \int dk' W(x, k') f(x, k + k', t), \quad (2a)$$

where

$$W(x, k') = \int dx' \sin(k' x') \times \left[V\left(x + \frac{x'}{2}\right) - V\left(x - \frac{x'}{2}\right) \right]. \quad (2b)$$

and where V is the (barrier) potential in the system. This term is added to the EMC transport equation and is used to update the particle distribution. The NLP term incorporates all quantum mechanical effects into the system. Apart from the NLP term, the remainder of the simulation is a standard EMC simulation, where the ensemble value is redefined as

$$\langle Q \rangle = \frac{\sum_i A(i) Q(i)}{\sum_i A(i)}, \quad (3)$$

where $A(i)$ is again the particle affinity and $Q(i)$ is the specific quantity of the system of interest (such as velocity or energy). As can be seen, if $A(i)$ is set to one, the regular definition of the ensemble is regained. Also, it is important to note that all particles in the system are treated normally as they would be in a standard EMC. That is, all particles in the system, regardless of affinity are drifted using the standard drift term, and all particles will be scattered, although scattering has not yet been added to the simulation. A full description of the method can be found in Shifren and Ferry (2001, in press). The method has been used to study a Gaussian wave-packet incident on a single potential barrier and to study a resonant tunneling diode (RTD).

Initially we were interested in the study of a Gaussian wave-packet incident on a single potential barrier, as the problem is well understood. Due to the fact that the problem is strictly quantum mechanical (that is, in the absence of any dissipation, the barrier is the only source of perturbation in the system), it gives a

clear indication of the effectiveness of including tunneling, correlation and interference into a EMC using our WDF approach. To fully test the effectiveness of the technique, the solutions to this problem were tested against solutions of a direct WDF solution and that of solving the Schrödinger equation. The results were not only evaluated by comparing resulting densities and transmission coefficients, but also by the use of Bohm trajectories. The results confirmed that our quantum EMC not only correctly calculated tunneling coefficients but also produced the correct Bohm trajectories (comparable to the other two fully quantum mechanical approaches). These results can be seen in Shifren and Ferry (2001). However it is interesting to view the actual phase-space distributions of the Gaussian as it interacts with the barrier to fully identify not only the interference and tunneling, but also the correlation that is naturally incorporated into the system. Figure 1 shows the Gaussian during its peak interaction with the barrier, that is, before the reflected and transmitted pulse has fully formed. It is important to note the large build up of "negative" density in the Wigner distribution before the barrier. This large negative region is thought to be a region of large uncertainty where no electrons may exist, and is a purely quantum mechanical phenomenon seen in our particle based EMC. Figure 2 is the same Gaussian however further along in the process when the transmitted and reflected wave-packets have fully formed. Here, the large correlation still exists between the two waves. As the waves move further apart the correlation between them will

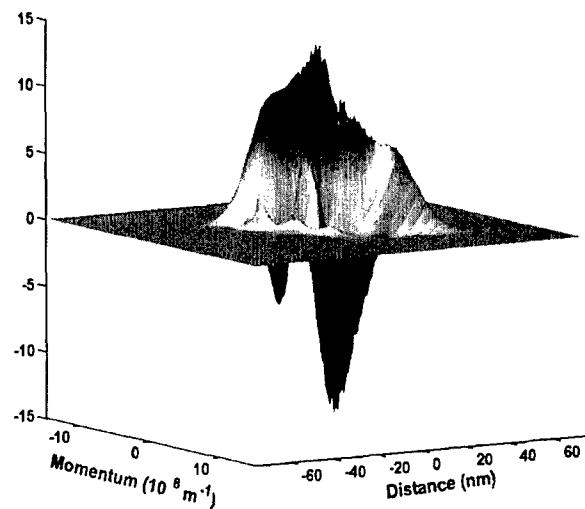


Figure 1. Gaussian wave-packet as it initially begins to interact with a single potential barrier which is 3 nm wide and 0.3 eV high.

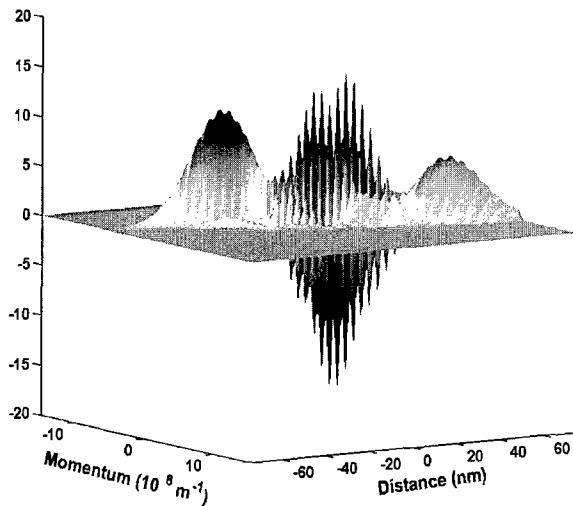


Figure 2. Gaussian wave-packet after it has interact with a single potential barrier, which is 3 nm wide and 0.3 eV high, and the transmitted and reflected wave-packets have been fully formed. The correlations between these two wave-packets increases as they move further apart.

continue to grow. The correlation allows the two waves to recombine and under time reversibility in the system. However, any form of dissipation in the system will destroy this correlation. It is clearly evident that the system correctly accounts for quantum effects, and more realistic devices and situations may also be studied.

The first device we have studied is the RTD, made up of GaAs bulk regions that are doped $1 \times 10^{18} \text{ cm}^{-3}$ with 3 nm AlGaAs intrinsic barrier regions and a 5 nm

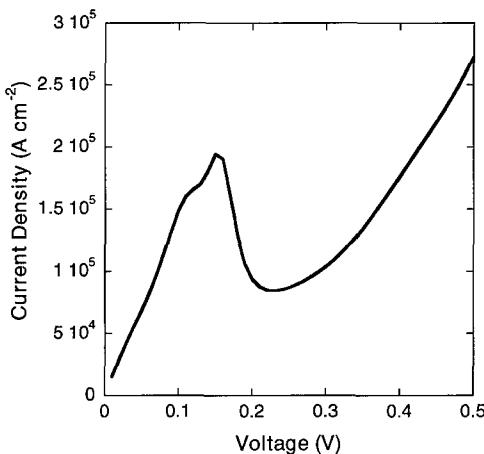


Figure 3. I-V curve for an RTD generated using the Wigner function quantum Monte Carlo.

GaAs intrinsic well region. The simulations were run unbiased for a few pico-seconds until steady-state was reached and this steady-state distribution was used as the initial state to compute I-V characteristics for the device. The simulation was run in incremental biases until steady-state was reached from 0 V to 0.5 V. The resulting I-V is seen in Fig. 3. Two things are evident. From Fig. 3 the expected RTD I-V characteristics are seen. The NDR is seen due to the resonant level which exists in the device being swept through as a bias is applied. If we consider Fig. 4, which is the phase-space distribution function located at the peak bias point of

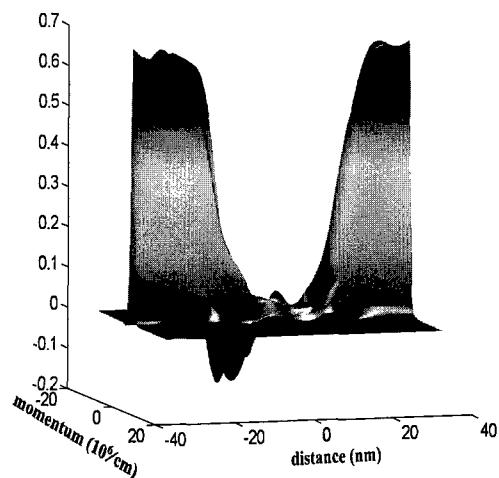


Figure 4. Phase-space distribution from the Wigner function quantum Monte Carlo generated at the peak bias point seen in Fig. 3.

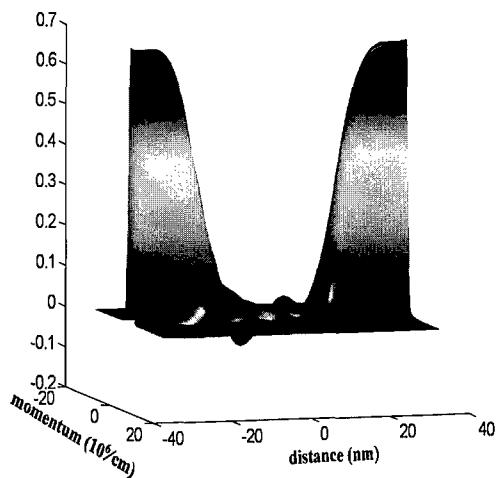


Figure 5. Phase-space distribution from the Wigner function quantum Monte Carlo generated at the valley bias point seen in Fig. 3.

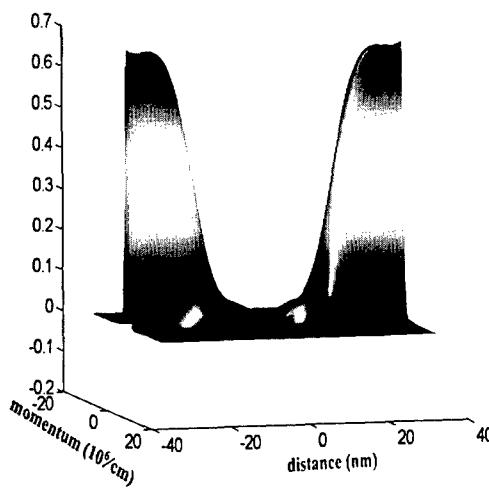


Figure 6. Phase-space distribution from the Wigner function quantum Monte Carlo generated at the maximum bias point seen in Fig. 3.

the I-V curve in Fig. 3, we notice that there is a large negative correlation that exists. As the device is resonant at the peak, it experiences an increase in tunneling through the barrier region which in turn generates large correlation between the transmitted and reflected densities. If the bias is increased to the valley of the I-V curve, the distribution of Fig. 5 is obtained. As can be seen, there is no longer a large amount of negative correlation, although it does still exist. The device is no longer in resonance and returns to the "normal" tunneling regime. Finally, Fig. 6 shows the distribution at the maximum bias point on the I-V curve. What is important to note here is that there is little if no correlation. The device is biased such that the 0.5 V applied is larger than the 0.3 eV barriers. As may be inferred from the figure the density flows over the barriers and is no longer in the tunneling regime.

By developing this new method of EMC, we hope to make it possible to include quantum mechanical

effects, namely tunneling, interference and correlation into an EMC simulation of more realistic devices, including full dissipation via non-local phonon scattering and self-consistency. EMC has long been the method of choice for device simulation due to its reliability and predictive capabilities. By correctly including quantum mechanical effects, this new development should lead to new approaches in the simulation and understanding of ultra-small devices.

Acknowledgments

This work is supported by the Office of Naval Research. The authors have enjoyed fruitful discussion with Mihail Nedjalkov and Christian Ringhofer.

References

- Akis R., Shifren L., Ferry D.K., and Vasileska D. 2001. *Phys. Stat. Sol. (b)* 226: 1.
- Ferry D.K. 1985. Granular nanostructures. In: Ferry D.K., Barker J.R., and C. Jacoboni (Eds.), New York: Plenum, pp. 1–18.
- Ferry D.K. and Grubin H.L. 1995. *Sol. State Phys.* 49: 283.
- Ferry D.K., Ramey S.M., Shifren L., and Akis R. This proceedings.
- Frensel W.R. 1987. *Phys. Rev. B* 36: 1570.
- Garcia-Garcia J., Martin F., Oriols X., and Sune J. 1998. *Appl. Phys. Lett.* 73: 3539.
- Jacoboni C., Bertoni A., Bordone P., and Brunetti R. 2001. *Math. Comp. Simul.* 55: 67.
- Kawaura H., Sakamoto T., Baba T., Ochiai Y., Fujita J., Matsui S., and Sone J. 1997. *Jpn. J. Appl. Phys.* 36: 1569.
- Kluksdahl N.C., Kriman A.M., Ferry D.K., and Ringhofer C. 1989. *Phys. Rev. B* 39: 7720.
- Ravaoli U., Osman M.A., Pötz W., Kluksdahl N.C., and Ferry D.K. 1985. *Physica B + C*, 134B: 36; Kluksdahl N.C., Pötz W., Ravaoli U., and Ferry D.K. 1987. *Supperlatt. Microstruct.* 3: 41.
- Shifren L., Akis R., and Ferry D.K. 2000. *Phys. Lett. A* 274: 75.
- Shifren L. and Ferry D.K. 2001. *Phys. Lett. A* 285: 217.
- Shifren L. and Ferry D.K. *Physica B*, in press.
- Wigner E. 1932. *Phys. Rev.* 40: 749.



The Effective Potential in Device Modeling: The Good, the Bad and the Ugly

D.K. FERRY, S. RAMEY, L. SHIFREN AND R. AKIS

Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University,
Tempe, AZ 85287-5706, USA

Abstract. We discuss the use of the effective potential to incorporate quantum effects in device models. While threshold shifts and charge set-back are handled well, tunneling is not well handled by this approach, or by any other local potential approach.

Keywords: device modeling, transport, quantization, tunneling

1. Introduction

Quantum effects are known to occur in the channel of MOSFETs, where the confinement is in the direction normal to the oxide interface. For quite some time, there has been a desire to categorize this quantization and determine the role it plays in semiconductor devices. Often, this is found by solving the Schrödinger and Poisson equations to find the actual position of the charge and the changes in mobility and capacitance (Vasileska *et al.* 1997). More recently, it has become of interest to include a *quantum potential* as a correction to the solutions of the Poisson equation in self-consistent simulations (Zhou and Ferry 1992). This latter approach has come to be called the “density-gradient” approach, since the quantum potential is defined in terms of the second derivative of the square root of local density. Such an approach is highly sensitive to noise in the local carrier density, and the methodology is highly suspect in cases of strong quantization (Ferry and Barker 1998).

We have developed a different approach, which introduces an *effective* potential. Here, the natural non-zero size of an electron wave packet in the quantized system, is used to introduce a smoothing of the local potential (found from Poisson’s equation) (Ferry 2001). This approach naturally incorporates the quantum potentials, which are *approximations* to the effective potential. The introduction of an effective potential follows two trends that have been prominent in statistical physics during most of the twentieth century and

into the current century. These are the non-zero size of an electron wave packet and the use of a modified potential to describe quantum effects within classical statistical mechanics. Here, we review these two approaches and show how they combine to give a form for the effective potential. We then show how the quantum potential derives from the effective potential as an approximation, and finally provide results from simulations to compare these approaches. We also estimate the problems in incorporating tunneling via this approach.

2. The Effective Potential

In order to describe the packet in real space, one must account for the contributions to the wave packet from all occupied plane wave states (Ferry 1998). That is, the states that exist in momentum space are the Fourier components of the real-space wave packet. If we want to estimate the size of this wave packet, we must utilize all Fourier components, not just a select few. (This approach is familiar from the definition of Wannier functions and their use to evaluate the size of a bound electron orbit near an impurity.) This is not the first attempt to define the nature of the quantum wave packet corresponding to a (semi-)classical electron. Indeed, the study of the classical-quantum correspondence has really intensified over the past few decades, due in no small part to the rich nature of chaos in classical systems and the search for the quantum analog of this

chaos. This has led to a number of studies of the manifestation of classical phase-space structure (Skodje *et al.* 1989). These have shown that meaningful sharp structure can exist in quantum phase-space representations, and these can profitably be used to explain (or to interpret) quantum dynamics; e.g. to study the quantum effects that arise in otherwise classical simulations for semiconductor devices. The use of a Gaussian wave packet as a representation of the classical particle is the basis of the well-known coherent-state representation. In the latter approach, the phase-space representation of the quantum density localized at point \mathbf{x} is given by Glauber (1963), Klauder (1963, 1964) and Klauder and Sudarshan (1968)

$$\langle \mathbf{x} | \mathbf{p}, \mathbf{q} \rangle = \frac{1}{(\pi\sigma^2)^{N/4}} \exp\left[-\frac{(\mathbf{x}-\mathbf{q})^2}{2\sigma^2} + i\frac{\mathbf{p} \cdot (\mathbf{x}-\mathbf{q})}{2\hbar}\right]. \quad (1)$$

As in most cases, the problem is to find the value of the spatial spread of the wave packet, which is defined by the parameter σ , which is related to the width of the wave packet.

At the same time, there has been a growing interest in methods which allow the reduction of quantum calculations to classical ones, through the introduction of a suitable *effective potential*. The earliest known approach was provided by Wigner (1932), where he introduced an expansion of the classical potential in powers of \hbar and $\beta = 1/k_B T$, which led to

$$V_{\text{eff}}(x) \sim V(x) + \frac{\hbar^2}{8mk_B T} \frac{\partial^2 V}{\partial x^2} + \dots \quad (2)$$

This series led to the well-known Wigner-Kirkwood expansion of the potential that is often used in solutions for the Wigner distribution function. However, the series has convergence problems below the Debye temperature and in cases with sharp potentials, such as the Si-SiO₂ interface. Feynman and Hibbs (1965) found a similar result, but with the factor 8 replaced by 24. He also introduced a different approach, in which an effective potential is introduced through the free energy. For the case of a free particle, he shows that the exact variational minimization leads to a Gaussian weighting of the potential around the classical path, and this automatically includes quantum effects into the trajectory. Indeed, Feynman found that the smoothing parameter σ should have the value

$$\sigma^2 = \frac{\hbar^2}{12mk_B T} = \frac{\lambda_D^2}{24\pi}, \quad (3)$$

where λ_D is the thermal de Broglie wavelength. The connection of this to our wave packet lies in the fact that the total Hamiltonian for a spatially varying potential involves weighting the potential at x by the density at this position. Then, the Gaussian spread of the density is easily transformed into a Gaussian weighting of the effective potential (Ferry 2001).

Many people have extended the Feynman approach to the case of bound particles (Giachetti and Tognetti 1985, Feynman and Kleinert 1986, Cao and Berne 1990, Voth 1991, Cuccoli *et al.* 1992) and particles at interfaces (Kriman and Ferry 1989). The effective potential approach has been recently reviewed by Cuccoli *et al.* (1995). These approaches use the fact that the most-likely trajectory in the path integral no longer follows the classical path when the electron is bound inside a potential well. The introduction of the effective potential and its effective Hamiltonian is closely connected to the return to a phase-space description, as discussed above. This can be done at present only for Hamiltonians containing a *kinetic* energy quadratic in the momenta and a coordinate-only dependence in the potential energy. That is, it is clear that some modifications will have to be made when non-parabolic energy bands, or a magnetic field, are present. However, the Gaussian approximation is well established as the method for incorporating the purely quantum fluctuations around the resulting path. The key new ingredient for bound states (such as in the potential well at the interface of a MOSFET) is the need to determine variationally the dominant path and hence the “correct” value for the parameter σ . For the case in which the bound states are well defined in the potential, both Feynman and Kleinert (1986) and Cuccoli *et al.* (1992) find

$$\sigma^2 = \frac{\hbar^2}{4mk_B T} \left[\frac{\coth(f)}{f} - \frac{1}{f^2} \right], \quad (4)$$

where

$$f = \frac{\hbar\omega_0}{2k_B T} \quad (5)$$

and $\hbar\omega_0$ is the spacing of the subbands. If we take the high-temperature limit, then we can expand for small f , and

$$\sigma^2 \sim \frac{\hbar^2}{12mk_B T} \quad (6)$$

to leading order, which agrees with (3). In Si, this gives a value of 0.52 nm for the value to be used in the

direction normal to the interface (at room temperature). A different mass would be used for transport along the channel, and this gives a value of 1.14 nm.

It is important to note that the density-gradient potential is easily derived as a low-order expansion to the actual effective potential, although there will be differences in the numerical factors among different approaches to this quantity (Ferry 2001). We can expand the effective potential when it is a slowly varying function of position. That is, we take the effective potential from the defining lines in (1) and use a Taylor series expansion as

$$\begin{aligned} W(x) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} V(x + \xi) e^{-\xi^2/2\sigma^2} d\xi \\ &\approx \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left[V(x) + \xi \frac{\partial V}{\partial x} \right. \\ &\quad \left. + \frac{\xi^2}{2} \frac{\partial^2 V}{\partial x^2} + \dots \right] e^{-\xi^2/2\sigma^2} d\xi. \end{aligned} \quad (7)$$

The first term allows us to bring the potential outside the integral, while the second term vanishes due to the symmetry of the Gaussian. The third term becomes the leading correction term, which gives us

$$V_{\text{eff}}(x) = V(x) + \sigma^2 \frac{\partial^2 V}{\partial x^2} + \dots \quad (8)$$

We note that this result gives a value for the smoothing parameter, if we compare with the results of Wigner (1932), of

$$\sigma^2 = \frac{\hbar^2}{8mk_B T} = \frac{\lambda_D^2}{16\pi}, \quad (9)$$

and a factor of 1.5 smaller for the Feynman result.

3. The Good

We may easily incorporate the *effective* potential into MOSFET simulations, as the Gaussian weighting is simply a multi-dimensional smoothing of the potential, which is found from Poisson's equation. A simulation, in which the transport is handled by an ensemble Monte Carlo approach, quite generally finds that the threshold voltage is shifted and the carrier density is moved away from the interface. Both effects are a result of quantization within the channel (Ferry *et al.* 2000). Treatment of an SOI device is discussed in a separate paper in this proceedings (Ramey and Ferry 2002), as is the role of

surface roughness scattering in the transient response of a MOSFET (Formicone *et al.* 2002). With the proper evaluation of the smoothing parameter σ , agreement with both the quantization energy and the amount of charge set-back from the interface are found to agree well with a full Poisson-Schrödinger simulation (Ferry 2001).

The above results show us that the effective potential is a very good approach in which to incorporate the quantum effects into device simulation. But, how does this approach compare with the density gradient approach? In order to answer this question, we have collaborated with the device simulation group at the University of Glasgow, headed by Prof. Asen Asenov, to simulate a simple MOSFET. Recently, the results of this collaboration were presented at SISPAD (Watling *et al.* in press). In this work, the quantum influence on threshold voltage, carrier density profile and I_D-V_G current characteristics were investigated within a modified drift diffusion framework. Results from the new effective potential algorithm were compared with those from the well-established density gradient approach. Here, it was found that the density-gradient approach agreed better with the Poisson-Schrödinger simulations, and that the effective potential pushed the charge too deep into the channel. However, the value of σ used was 0.7 nm, which is 40% too large and probably accounts for these results. One must be careful here, as the exact values of the smoothing parameters used will dramatically effect the position and value of the peak in the density, as will the grid spacing used.

We have also shown that by using an appropriate effective potential, obtained by convolving the self-consistent potential with a Gaussian, we can replicate certain quantum behavior in a quantum point contact by using classical physics (Shifren *et al.* 2000). Significantly, in contrast to the Bohm potential method, one is not required to actually solve Schrödinger's equation in all situations using this method. While densities entering into the Poisson equation were obtained quantum mechanically in this study (necessitated by the strong quantization in this particular quasi-one-dimensional system), one can obtain good results simply by convolving the potentials obtained from a particle-based Poisson solver.

4. The Bad

Various forms for the *quantum* (e.g., the density-gradient) potential are really approximations to the full

effective potential. As a result, use of the latter is to be preferred, since the integral smoothing will reduce fluctuations while the derivative forms amplify fluctuations. Moreover, the effective potential carries the entire quantization effects, which arise from the non-zero size of the electron wave packet. This means that the effective potential is already of a nature to be used for mixed wave functions, whereas the density gradient approaches have severe problems in this case, particularly near nodal points of the composite wave function (Ferry and Barker 1998). The problem, however, is in deciding upon the size of the smoothing parameter σ . As mentioned above, a value near 0.5 nm is believed to be correct based upon an evaluation of the bound energy levels in the quantum well formed at the Si-SiO₂ interface. The value for motion along the interface, however, is not so well determined.

Feynman found that the smoothing parameter σ should have the value given by (3) for free particles. On the other hand, a different result can be found by taking the approach of Wannier functions. In the ideal case, the Wannier function is an atomically-sharp, localized wave packet formed by a sum over all Bloch functions within a (full) band. However, if the band is only partially full, as is the case in semiconductors, then the sum should only run over the occupied states. Assuming that these are given by a Maxwell-Boltzmann distribution, we are led to a value of the smoothing parameter of Ferry (2001)

$$\sigma^2 = \frac{3\lambda_D^2}{16}, \quad (10)$$

which gives a value for σ of 16.1 nm, which is some 14 times larger than the 1.14 nm obtained from (3). The result from Wigner (1932) gives a value of 1.7 nm, which lies in between these two limits. Thus, different theoretical approaches give a rather large degree of uncertainty in this value.

The importance of the smoothing parameter, for transport along the channel, lies in its effect on the source-channel barrier, which governs transport in the MOSFET. Over-smoothing of this barrier reduces its effective height too much, which results in an overly large drain-induced barrier lowering (DIBL) and affects the source-drain tunneling. While we can gain a quite good estimate of the correct value for the normal direction from coupled Poisson-Schrödinger solutions, this is not the case for the value along the channel. Here, there is usually insufficient information on the details of the source, channel, and extension dopings that are

present in actual devices to use either measurements or simulations of DIBL to unfold the best value for σ . Source-drain tunneling is also not a good test, for there is considerable doubt over the ability of effective potentials to correctly simulate tunneling processes (discussed below). However, the use of an effective potential greatly affects the resulting DIBL, particularly in sub-threshold situations, that is found in simulations. As a consequence, it is important to get better data on actual fabricated ultrasmall devices, particularly the actual values of the various impurity concentrations, sidewall spacers, and oxide thicknesses (including the transitional SiO_x regions).

An alternative approach, which may shed light on the proper values to take for the smoothing parameter, may be found from studies of mesoscopic structures fabricated in Si. For example, recent studies of Si quantum dots could be used to study carefully the potential barriers, and the transmission through these barriers, as a means of evaluating the smoothing introduced by the effective potential. While quantum point contacts introduce lateral confinement potentials, it is the shape in the longitudinal part of the saddle potential that affects transport. We have shown that classical particles can be induced to follow quantum behavior in such quantum point contacts (Shifren *et al.* 2000), so the study of these could shed light on the proper values to use in MOSFET simulations. It is hoped that such studies will appear in the not-too-distant future.

5. The Ugly—Tunneling

It is absolutely clear that a hydrodynamic approach to the solution of quantum tunneling has given effective results. This was first demonstrated by Dewdney and Hiley (1982), when they solved the Schrödinger equation and used the quantum potential to determine the trajectories flowing through the barrier (and those that were reflected). This has been repeated even for studies of quantum chemistry (Wyatt 1999). In general, these studies use a Gaussian wave packet, which impinges upon a barrier. However, it has also been shown that the shape of the packet itself affects the tunneling coefficient (Lopreore and Wyatt 1999). The trajectories that make it “over” the barrier are first accelerated by the quantum potential that exists at the initial time of the simulation. Thus, a sharper packet will give more tunneling, but it also contains more high momentum states that can actually go over the barrier. It is important to remember here that these simulations actually solve the

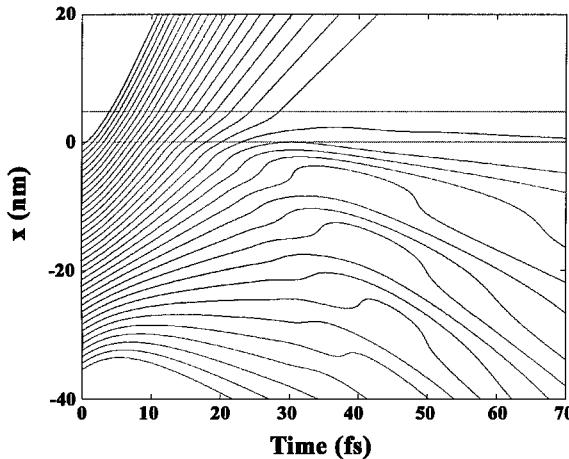


Figure 1. Bohm trajectories for a wave packet tunneling through a barrier located between the two horizontal lines. Only trajectories starting in the front of the wave packet (higher momentum states) have sufficient momentum to traverse the barrier.

Schrödinger equation, with an ex post facto determination of the quantum potential, even when solved in a density-velocity space. Such a solution for the resulting trajectories is shown in Fig. 1, and similar solutions have been obtained with Wigner function simulations (Shifren and Ferry 2001).

However, when we attempt to solve for a tunneling problem using only the density-gradient approach, without solving the Schrödinger equation, we do not get good results. One can consider just why this occurs. While the density is continuous at the hetero-interface between the semiconductor channel and the tunneling barrier (continuity of the wave function), its derivative is not necessarily continuous. In fact, the density must decay quite rapidly due to the fact that this state lies well below the barrier energy. This gives a sharp spike arising from the second derivative of the density, so that the quantum potential is composed of significant discontinuities, and a smooth tunneling behavior is not achieved. In fact, in our own simulations of this, we find significant charge storage in the barrier, which is not accounted for in any classical or semi-classical approach. This charge storage is thought to be non-physical and a result of the inapplicability of the density-gradient approach to tunneling problems. The failure can be traced to a deeper physical meaning, and that is the fact that tunneling occurs in quantum mechanics through a non-local effect of the barrier on the wave function. This non-locality is easily seen when a phase-space representation, such as the Wigner function, is used. When we use the density-gradient potential, or even the

effective potential, we are changing the problem to a local one, which means that we should not see quantum effects such as tunneling.

The failure in tunneling problems is also present with the effective potential. We can study this to some extent by using a Gaussian wave packet and propagating it through the tunneling barrier using the Wigner equation of motion. We use a barrier height of 0.3 eV and thickness of 3 nm. However, instead of using the full non-local potential terms, we replace these with a smoothed local potential. There are two ways to approach this. In the first, we use a series of wave packets with different spatial extents while keeping the smoothing of the barrier at 0.5 nm. Then, for wave packets with $\sigma = 0.5, 2, 4$, and 8 nm, the same tunneling coefficient $T = 0.08$ is obtained. The only problem is that the tunneling coefficient expected for the mean momentum of the wave packets is 0.15!

In the second approach, we smooth the barrier with a Gaussian of the same width as the incoming wave packet. These results are shown in Fig. 2, where we plot the tunneling coefficient as a function of the energy corresponding to the mean momentum of the phase space wave packet. Here, the tunneling coefficient varies from 0.02 to large values for packets whose energy varies from 0.12 eV to >0.2 eV, respectively. Again, however, the computed tunneling coefficient for the mean

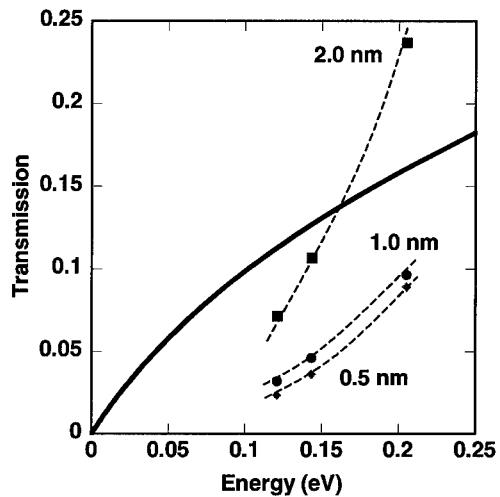


Figure 2. The tunneling coefficient for a wave packet whose standard deviation is used for the smoothing Gaussian of the potential. Here we plot the tunneling coefficient as a function of the energy corresponding to the mean momentum of the phase space wave packet. The solid curve is that for a single plane wave of the same energy, without any smoothing of the barrier. The width of the packet is a parameter on the calculated curves.

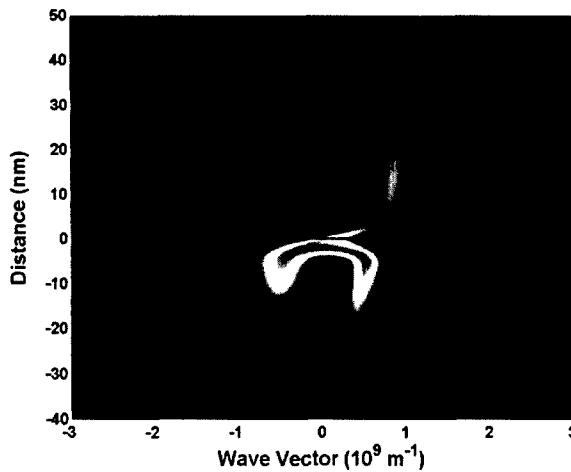


Figure 3. A wave packet is interacting with the barrier (delineated by the two dark lines). Note that, while most of the packet is reflected, the transmitting part is actually accelerated during passage through the barrier region. This is thought to be unphysical, and is a result of numerical diffusion in momentum space.

momentum of the packets is shown by the solid curve, and there is no agreement in the results. In essence, the “tunneling coefficient” is largely a result of significant fractions of the wave packet undergoing classical transport over the barrier, which we show in Fig. 3.

One can think about the width of the packet inducing some trajectories, with momentum corresponding to energies above the barrier, to pass the barrier classically. However, wider packets have fewer of these trajectories, but they are still important. Thus, we express the results in terms of “over the barrier” effects. This is coupled to the fact that the increased smoothing is actually *lowering* the barrier height, as shown in Fig. 4, so that the tunneling coefficient for the smoothed barrier increases. The tunneling behavior exhibited by these approximations, using both the density gradient potential and the effective potential, do not give meaningful results for tunneling. On the other hand, Ancona *et al.* (2000) claim to be able to treat tunneling with the density gradient approach. Notwithstanding the fact that improper boundary conditions are used in the simulation (one cannot have both the density and the quasi-Fermi energy continuous at the hetero-interface), the results are interesting but are more likely a result of having too many unknown parameters with which to play. More work needs to be done on such simulations to ascertain whether the above arguments preclude achieving good simulations of tunneling, in which case the latter results are merely fortuitous, or whether either of these approaches can be amended to

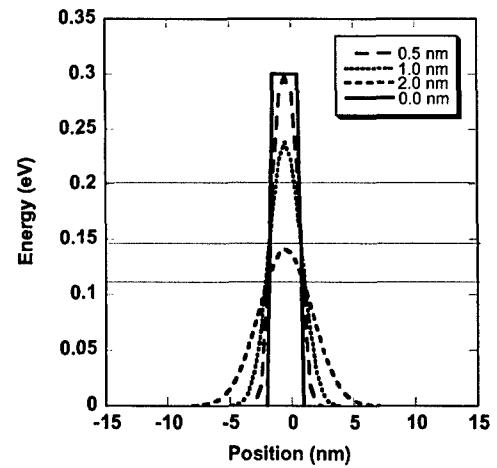


Figure 4. The effective potential barrier after smoothing allows the average energy of the wave packet to approach, or even surpass, the peak of the barrier. The horizontal lines correspond to mean momentum in the simulation for Fig. 2.

allow treatment of tunneling in device simulations, as claimed by Ancona *et al.*

6. Discussion

The effective potential approach has been successfully used to account for quantization effects in several simulations of a MOSFET. The effective potential provides a set-back of the charge from the interface, and a quantization energy within the channel. Both of these effects lead to an increase in the threshold voltage, which is apparent in the output characteristics of the device itself. However, the transport of the carriers along the channel, which is in a direction normal to the quantization direction, suffers from not having a good theoretical expression for the smoothing, and good experiments to clarify this have not yet been done. The approach using an effective potential automatically includes the density-gradient approach, which is at best an approximation to the more accurate effective potential. The computational cost of the effective potential approach is low, with less than a 10% increase in cpu time required to smooth the potential. As a result, this approach is readily incorporated within standard simulators at modest increase in complexity.

On the other hand, the use of these approximate potentials for tunneling problems counteracts the non-locality of tunneling itself by using some form of a local potential. As a result, really nasty results are often obtained from this approach, and it is not clear whether

or not some corrections can be added to improve the situation.

Acknowledgments

This work was supported by the Office of Naval Research, the National Science Foundation through the Descartes Center, and the Semiconductor Research Corporation. The authors have enjoyed many helpful discussions with J.R. Barker, A. Asenov, D. Vasileska, A. Demkov, D. Javonovic, R. Dutton, M. Ancona, and K. Smith.

References

- Ancona M. *et al.* 2000. IEEE Trans. Electron Dev. 47: 2310.
 Cao J. and Berne B.J. 1990. J. Chem. Phys. 92: 7531.
 Cuccoli A. *et al.* 1992. Phys. Rev. B 45: 2088.
 Cuccoli A. *et al.* 1995. J. Phys. Cond. Matter 7: 7891.
 Dewdney C. and Hiley B.J. 1982. Found. Phys. 12: 27.
 Ferry D.K. 1998. In Proc. IWCE, Kyoto, October 1998, IEEE Press.
 Ferry D.K. 2001. Superlatt. Microstruc. 27: 59. VLSI Design 13: 155.
 Ferry D.K. and Barker J.R. 1998. VLSI Design 8: 165.
 Ferry D.K. *et al.* 2000. Proc. IEDM, IEEE Press, New York, p. 287.
 Feynman R.P. and Kleinert H. 1986. Phys. Rev. A 34: 5080.
 Feynman R.P. and Hibbs A.R. 1965. Quantum Mechanics and Path Integrals. McGraw-Hill, New York.
 Formicone G.F. *et al.* 2002. J. Comp. Electron. 1: 1–2.
 Giachetti R. and Tognetti V. 1985. Phys. Rev. Lett. 55: 912.
 Glauber R.J. 1963. Phys. Rev. 131: 2766.
 Klauder J.R. 1963/1964. J. Math. Phys. 4: 1055; 4: 1058; 5: 177.
 Klauder J.R. and Sudarshan E.C.G. 1968. Fundamentals of Quantum Optics. Benjamin, New York.
 Kriman A. and Ferry D.K. 1989. Phys. Lett. A 138: 8.
 Lopreore C.L. and Wyatt R.E. 1999. Phys. Rev. Lett. 82: 5190.
 Ramey S.M. and Ferry D.K. 2002. J. Comp. Electron. 1: 1–2.
 Shifren L. and Ferry D.K. 2001. Phys. Lett. A 285: 217.
 Shifren L. *et al.* 2000. Phys. Lett. A 274: 75.
 Skodje R.T. *et al.* 1989. Phys. Rev. A 40: 2894, and references therein.
 Vasileska D. *et al.* 1997. IEEE Trans. Electron Dev. 44: 577; 44: 584.
 Voth G.A. 1991. J. Chem. Phys. 94: 4095.
 Watling J.R. *et al.* Proc. SISPAD, in press.
 Wigner E.P. 1932. Phys. Rev. 40: 749.
 Wyatt R.E. 1999. J. Chem. Phys. 111: 4406.
 Zhou J.R. and Ferry D.K. 1992. IEEE Trans. Electron Dev. 39: 473.



Wigner Paths for Quantum Transport

PAOLO BORDONE AND CARLO JACOBONI

*INFM and Dipartimento di Fisica, Università di Modena e Reggio Emilia, Via Campi 213/A,
I-41100 Modena, Italy*

Abstract. A Monte Carlo algorithm based on the concept of Wigner paths has been developed to study quantum transport in mesoscopic systems in strict analogy with the traditional Monte Carlo simulation used to solve the Boltzmann transport equation. Scatterings with both phonons and impurities can be accounted for. As regards a structure potential profile the effect of the corresponding classical force can be inserted in the dynamics of the free flight, while quantum effects due to rapid potential variations are included as a special scattering mechanism.

Keywords: Wigner paths, quantum transport, Monte Carlo, mesoscopic systems

1. Introduction

The Wigner-path (WP) concept, developed by the group of the authors in the recent years (Pascoli *et al.* 1998, Bertoni *et al.* 1999, Jacoboni *et al.* 2001), is based on the linearity of the dynamical equation for the Wigner function (WF). WP's are defined as the paths followed by "simulative particles" carrying δ -contributions of the WF through the Wigner phase-space, and are formed by ballistic free flights separated by scattering processes. Scattering with phonons, impurities, and an arbitrary potential profile can be included. Thus, the integral transport equation can be solved by a Monte Carlo (MC) technique by means of simulative particles following classical trajectories, in complete analogy to the "Weighted Monte Carlo" solution of the Boltzmann equation in the integral form. More precisely, the solution of the Wigner equation is obtained as a sum of contributions calculated along WP's formed by ballistic fragments, described by classical dynamics, separated by interaction vertices due to electron-phonon or potential interactions. The authors have developed a MC code based on the above concepts, in strict analogy with the traditional MC simulation technique used to study semi-classical transport phenomena.

2. The Physical System

The general system we are considering is formed by one electron (or, equivalently, many non-interacting electrons) subject to a constant and uniform accelerating field \mathbf{E} , to a structure potential (or to a given configuration of imputities (Menziani, Rossi and Jacoboni 1989)) $V(\mathbf{r})$, and to the interaction with phonons. The Hamiltonian of the system is given by

$$\mathbf{H} = \mathbf{H}_0 + V(\mathbf{r}) + V_f(\mathbf{r}) + \mathbf{H}_p + \mathbf{H}_{e-p}, \quad (1)$$

where

$$V_f(\mathbf{r}) = -e\mathbf{E} \cdot \mathbf{r},$$

and

$$\mathbf{H}_0 = -\frac{\hbar^2}{2m} \nabla^2,$$

$$\mathbf{H}_p = \sum_q \mathbf{b}_q^\dagger \mathbf{b}_q \hbar \omega_q,$$

$$\mathbf{H}_{e-p} = \sum_q i\hbar F(q) (\mathbf{b}_q e^{iq\mathbf{r}} - \mathbf{b}_q^\dagger e^{-iq\mathbf{r}}),$$

are the free electron term (with m electron effective mass), the Hamiltonian for the free phonon system

and the electron-phonon interaction term, respectively. In the above expressions \mathbf{b}_q and \mathbf{b}_q^\dagger are the annihilation and creation operators for the phonon mode \mathbf{q} , ω_q is the frequency of the phonon mode \mathbf{q} , and $F(\mathbf{q})$ is a function depending on the type of phonon scattering analyzed. The generalized WF (Rossi, Jacoboni and Nedjalkov 1994) for an electron-phonon system is:

$$\begin{aligned} f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) \\ = \int d\mathbf{r}' e^{-i\mathbf{p}\cdot\mathbf{r}'/\hbar} (\mathbf{r} + \mathbf{r}'/2, \{n_q\}) |\rho(t)| (\mathbf{r} - \mathbf{r}'/2, \{n'_q\}) \end{aligned} \quad (2)$$

where n_q is the occupation number of the phonon mode \mathbf{q} (the curly brackets indicate a set of phonon occupation numbers for all possible phonon modes), and ρ is the density operator of the electron-phonon system. Tracing over the phonon coordinates provides the WF representative of the electron system (Brunetti, Jacoboni and Rossi 1989).

3. Integral Equation and Neumann Expansion

Taking the time derivative of Eq. (2) and using the Liouville-von Neumann equation for the evolution of the density matrix, we obtain:

$$\begin{aligned} \frac{\partial}{\partial t} f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) \\ = \frac{1}{i\hbar} \int ds e^{-i\hbar\mathbf{p}\cdot\mathbf{s}} \left(\mathbf{r} + \frac{\mathbf{s}}{2}, \{n_q\} | [\mathbf{H}, \rho(t)] | \mathbf{r} - \frac{\mathbf{s}}{2}, \{n'_q\} \right) \end{aligned} \quad (3)$$

Using the Hamiltonian given in Eq. (1) the r.h.s. of the above equation can be written as the sum of five terms. Developing the calculations (the full derivation is given in Bertoni *et al.* (1999) leads to

$$\begin{aligned} & \left(\frac{\partial}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla - e\mathbf{E} \cdot \nabla_{\mathbf{p}} \right) f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) \\ &= \frac{1}{i\hbar} (\mathcal{E}(\{n_q\}) - \mathcal{E}(\{n'_q\})) f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) \\ &+ \int d\mathbf{p}' \mathcal{V}_w(\mathbf{r}, \mathbf{p}' - \mathbf{p}) f_w(\mathbf{r}, \mathbf{p}', \{n_q\}, \{n'_q\}, t) \\ &+ (\mathcal{S}_{ph} f_w)(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t). \end{aligned} \quad (4)$$

where

$$\mathcal{E}(\{n_q\}) = \sum_q n_q \hbar \omega_q \quad (5)$$

is the energy of the phonon state $\{n_q\}$, the transfer function \mathcal{V}_w is defined by

$$\begin{aligned} \mathcal{V}_w(\mathbf{r}, \mathbf{p}) = \frac{1}{\hbar^3} \int ds e^{-i\hbar\mathbf{p}\cdot\mathbf{s}} \frac{1}{i\hbar} \left[V\left(\mathbf{r} + \frac{\mathbf{s}}{2}\right) \right. \\ \left. - V\left(\mathbf{r} - \frac{\mathbf{s}}{2}\right) \right], \end{aligned} \quad (6)$$

and

$$\begin{aligned} & (\mathcal{S}_{ph} f_w)(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) \\ &= \sum_{\mathbf{q}'} F(\mathbf{q}') \left\{ e^{i\mathbf{q}'\cdot\mathbf{r}} \sqrt{n_{\mathbf{q}'} + 1} \right. \\ & \times f_w\left(\mathbf{r}, \mathbf{p} - \frac{\hbar\mathbf{q}'}{2}, \{n_1, \dots, n_{\mathbf{q}'} + 1, \dots\}, \{n'_q\}, t\right) \\ & - e^{-i\mathbf{q}'\cdot\mathbf{r}} \sqrt{n_{\mathbf{q}'}} \\ & \times f_w\left(\mathbf{r}, \mathbf{p} + \frac{\hbar\mathbf{q}'}{2}, \{n_1, \dots, n_{\mathbf{q}'} - 1, \dots\}, \{n'_q\}, t\right) \\ & - e^{i\mathbf{q}'\cdot\mathbf{r}} \sqrt{n'_{\mathbf{q}'}} \\ & \times f_w\left(\mathbf{r}, \mathbf{p} + \frac{\hbar\mathbf{q}'}{2}, \{n_q\}, \{n'_1, \dots, n'_{\mathbf{q}'} - 1, \dots\}, t\right) \\ & + e^{-i\mathbf{q}'\cdot\mathbf{r}} \sqrt{n'_{\mathbf{q}'} + 1} \\ & \times f_w\left(\mathbf{r}, \mathbf{p} - \frac{\hbar\mathbf{q}'}{2}, \{n_q\}, \{n'_1, \dots, n'_{\mathbf{q}'} + 1, \dots\}, t\right) \left. \right\}, \end{aligned} \quad (7)$$

is the contribution of the electron-phonon interaction. Each term on the r.h.s of Eq. (7) represents a phonon interaction event (vertex) that changes only one set of phonon coordinates, increasing or decreasing the phonon occupation number of mode \mathbf{q}' by one unity and changing the electron momentum by $\frac{\hbar\mathbf{q}'}{2}$. The l.h.s. of Eq. (4) has the same form as the classical Boltzmann equation (BE). Thus path variables can be used in analogy with the Chambers formulation of transport. Then, integrating over time, one obtains

$$\begin{aligned} & f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) \\ &= f_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t_s), \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; t_s), \{n_q\}, \{n'_q\}, t_s) \\ & \times e^{-(i/\hbar)(\mathcal{E}(\{n_q\}) - \mathcal{E}(\{n'_q\}))(t - t_s)} \end{aligned}$$

$$\begin{aligned}
& + \int_{t_0}^t dt' e^{-(i/\hbar)(\epsilon(\{n_q\}) - \epsilon(\{n'_q\}))(t-t')} \left\{ \int d\mathbf{p}' \right. \\
& \times \mathcal{V}_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \mathbf{p}' - \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; t')) \\
& \times f_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \mathbf{p}', \{n_q\}, \{n'_q\}, t') \\
& + (S_{ph}f_w)(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \\
& \quad \left. \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \{n_q\}, \{n'_q\}, t') \right\}. \quad (8)
\end{aligned}$$

$$\begin{cases} \mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; s) = \mathbf{r} - \frac{\mathbf{p}}{m}(t-s) + \frac{\mathbf{F}}{2m}(t-s)^2 \\ \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; s) = \mathbf{p} - \mathbf{F}(t-s) \end{cases} \quad (9)$$

are the position and the momentum of the particle at time s if at time t it has position \mathbf{r} and momentum \mathbf{p} and the force \mathbf{F} is constant; the upper (0) indicates that no scattering occurs between s and t . t_0 represents the time of the initial condition, when the WF is supposed to be known.

In more compact form, Eq. (8) can be written as

$$\begin{aligned} f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) = & f_0 + \mathbf{S}_V f + \mathbf{S}_{af} f + \mathbf{S}_{ef} f \\ & + \mathbf{S}_{as} f + \mathbf{S}_{es} f \end{aligned} \quad (10)$$

where

$$\begin{aligned} f_0 = & f_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t_0), \\ & \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; t_0), \{n_q\}, \{n'_q\}, t_0) \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{S}_V f = & \int_{t_0}^t dt' \int d\mathbf{p}' \mathcal{V}_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \\ & \mathbf{p}' - \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; t')) f_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \\ & \mathbf{p}', \{n_q\}, \{n'_q\}, t') \end{aligned} \quad (12)$$

$$\begin{aligned} \mathbf{S}_{af} f = & \int_{t_0}^t dt' \sum_{q'} F(\mathbf{q}') e^{i\mathbf{q}'\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t')} \sqrt{n_{q'} + 1} \\ & \times f_w(\mathbf{r}^{(0)}(\mathbf{r}, \mathbf{p}, t; t'), \mathbf{p}^{(0)}(\mathbf{r}, \mathbf{p}, t; t') \\ & - \hbar\mathbf{q}'/2, \{n_1, \dots, n_{q'} + 1, \dots\}, \{n'_q\}, t'), \end{aligned} \quad (13)$$

and $\mathbf{S}_{ef} f$, $\mathbf{S}_{as} f$, $\mathbf{S}_{es} f$ are defined, similarly to $\mathbf{S}_{af} f$, as the time integration of the three last terms on the r.h.s. of Eq. (7), respectively; a and e indicate an absorption and an emission event, respectively, while f and s refer to the first and second phonon arguments, as defined in Eq. (2).

Equation (10) may be iteratively substituted into itself giving a Neumann expansion that describes the

evolution of the WF as a sum of contributions containing increasing powers of the interaction coupling:

$$\begin{aligned} f_w(\mathbf{r}, \mathbf{p}, \{n_q\}, \{n'_q\}, t) = & f_0 + \sum_i \mathbf{S}_i f_0 + \sum_{i,j} \mathbf{S}_i \mathbf{S}_j f_0 \\ & + \sum_{i,j,k} \mathbf{S}_i \mathbf{S}_j \mathbf{S}_k f_0 + \dots \end{aligned} \quad (14)$$

where $i, j, k = V, af, ef, as, es$.

4. “Particle” Simulation and Wigner Paths

Equation (14) is expressed as a sum of terms, then can be evaluated by means of the MC technique (Rossi, Poli and Jacoboni 1992). Given the sum

$$\mathcal{S} = \sum_i a_i \quad (15)$$

a possible MC algorithm for its evaluation is the following: a set of arbitrary probabilities p_i are defined, subject to the conditions

$$p_i \geq 0 \quad (p_i > 0 \text{ if } a_i \neq 0), \quad i = 1, 2, \dots, \quad \sum_i p_i = 1. \quad (16)$$

Then a term a_i is selected with the probabilities p_i , and the estimator

$$s = \frac{a_i}{p_i} \quad (17)$$

is evaluated. This is a correct estimator of the sum \mathcal{S} , since its expectation value is

$$\langle s \rangle = \sum_i p_i \frac{a_i}{p_i} = \mathcal{S}. \quad (18)$$

If, instead of a single sum \mathcal{S} , we have to evaluate a set of sums

$$\mathcal{S}_k = \sum_i a_{ki} \quad (19)$$

a set of arbitrary probabilities p_{ki} are defined, subject to the conditions

$$p_{ki} \geq 0 \quad (p_{ki} > 0 \text{ if } a_{ki} \neq 0), \quad i = 1, 2, \dots, \quad \sum_{ki} p_{ki} = 1. \quad (20)$$

Then a term a_{ki} is selected with probability p_{ki} and the estimators

$$s_j = \frac{a_{ki}}{p_{ki}} \delta_{kj} \quad (21)$$

are evaluated, where δ_{kj} is the Kronecker symbol. These are correct estimators of the sums S_j ; in fact their expectation values are

$$\langle s_j \rangle = \sum_{ki} p_{ki} \frac{a_{ki}}{p_{ki}} \delta_{kj} = \sum_i a_{ji} = S_j. \quad (22)$$

It should be noticed that the selection of a single term of the matrix a_{kj} yields an estimate of all the sums in Eq. (19); this estimate is a_{ki}/p_{ki} for the k th sum and zero for the other sums.

The above algorithm can be easily generalized to evaluate integrals, and sum of integrals (Rossi, Poli and Jacoboni 1992). It should be emphasized that the probabilities used in the algorithm are arbitrary. The correctness of the estimator does not depend on them, while, on the other hand, its variance does. A suitable choice of the probabilities may reduce drastically the variance of the result.

Moving one step further, we notice that each term of Eq. (14) is, of its own, a sum, and again, each term of such sum contains further sums and integrals (see Eqs. (12) and (13)). The MC solution must select a particular term of Eq. (14), then the single contribution to the sum (that is the sequence $i j k \dots$), then the value of the sum appearing in the integrand, and finally the value of the integrand function. Finally, in the estimator, the particular selected value of the integrand is divided by the probability of that choice. This combination of choices corresponds to select the number of scattering processes, their sequence, the exchanged momenta, and the scattering times. Once the scattering times are determined the coordinates in the phase-space are related to the ones in the WF argument on the l.h.s. of Eq. (14) through relations of the type of those given in Eq. (9). Thus each term of the series can be treated as the weighted contribution of a path consisting of segments of classical trajectories separated by scattering events. This is a so called Wigner path (Pascoli *et al.* 1998, Bertoni *et al.* 1999, Jacoboni *et al.* to appear). Once the exact correspondence between a specific term (a_i for the case of Eq.(15), a_{ki} for the case of Eq. (19)) and a specific WP is understood, then it is clear that WP's can be chosen in different ways; as for example following the same procedure of the

traditional MC codes used for studying semiclassical transport.

If the argument of the WF on the l.h.s. of Eq. (14) is fixed the Neumann expansion is of the type of Eq. (15) where the various a_i are the possible paths starting at time t from the phase-space point (\mathbf{r}, \mathbf{p}) and going back in time up to the initial condition (backward procedure). On the other hand, we may leave the argument of the WF undefined. In this case the Neumann expansion is of the type of Eq. (19) and the WP can be chosen starting from a particular point of the phase-space and moving forward in time. The particular choice of the term a_{ki} leads to estimate the value a_{ki}/p_{ki} for the k th sum, corresponding to a particular value of the WF in the final point reached at the time t , while the contribution to the WF for all the other points in the phase-space has to be taken zero. Whatever approach is chosen, the solution of Eq. (14) is then obtained by generating a very large number of paths, until the required precision in the result is achieved.

5. Towards an Efficiency Increase

Since, as just mentioned, the above described MC approach is based on the simulation of a number of WP's, where the larger is the number of paths accounted for, the better is the statistical precision achieved, the main limitation to an extensive application of the method has been, so far, the simulation times required to obtain reliable results.

To improve the efficiency of the codes we have developed a method that allows to treat the scattering of the carrier with the potential profile separating the effect of the classical force from quantum corrections, and we have included in our algorithms the quantum self-scattering mechanism (Rossi and Jacoboni 1992).

It is well known that the effect of a potential $V(\mathbf{r})$ in the Wigner equation, given by the integral term of the type in Eq. (4), reduces to the classical-force term on the l.h.s. of the Wigner equation for potentials up to quadratic. In the general case, it is possible to separate the effect of the classical force from quantum corrections by defining the quantity

$$\tilde{V}(\mathbf{r}, \mathbf{r}') = V(\mathbf{r} + \mathbf{r}') - \nabla V(\mathbf{r}) \cdot \mathbf{r}'.$$

Now the dynamical equation for the Wigner function (see Eq. (4) where, for simplicity, phonon scattering

has been removed) becomes

$$\begin{aligned} \frac{\partial f_w}{\partial t} + \frac{\mathbf{p}}{m} \nabla f_w + \mathbf{F} \nabla_{\mathbf{p}} f_w(\mathbf{r}, \mathbf{p}) \\ = \frac{1}{(2\pi\hbar)^3} \int d\mathbf{p}' \tilde{\mathcal{V}}_w(\mathbf{r}, \mathbf{p} - \mathbf{p}') f_w(\mathbf{r}, \mathbf{p}'), \end{aligned} \quad (23)$$

where $\mathbf{F} = -(eE + \nabla V(\mathbf{r}))$ is the classical force, and $\tilde{\mathcal{V}}_w(\mathbf{r}, \mathbf{p})$, the usual integral kernel with \tilde{V} in place of V , is a term including only quantum corrections to the classical orbits. The l.h.s. of Eq. (23) is identical to the Liouvillian of the BE, while the first term on the r.h.s. describes quantum effects in the form of a collisional integral due to a sort of “quantum potential” (a similar approach was introduced by Lozovik and Filinov (1999)).

A further improvement in the efficiency of the algorithm should derive from the inclusion of the quantum self-scattering mechanism. This method is based on the introduction of an appropriate immaginary part of the self-energy $\Gamma = 1/\tau$ which plays a role analogous to that of the maximum scattering rate in the traditional MC method. Let us define

$$\tilde{f}_w(\mathbf{r}, \mathbf{p}, t) = e^{\Gamma(t-t_0)} f_w(\mathbf{r}, \mathbf{p}, t), \quad (24)$$

performing the derivative with respect to time we get

$$\frac{\partial}{\partial t} f_w = -\Gamma e^{-\Gamma(t-t_0)} \tilde{f}_w + e^{-\Gamma(t-t_0)} \frac{\partial}{\partial t} \tilde{f}_w. \quad (25)$$

Substituting Eq. (25) into Eq. (23) and using Eq. (24) leads to

$$\begin{aligned} \frac{\partial}{\partial t} \tilde{f}_w + \frac{\mathbf{p}}{m} \nabla \tilde{f}_w + \mathbf{F} \nabla_{\mathbf{p}} \tilde{f}_w \\ = \frac{1}{(2\pi\hbar)^3} \int d\mathbf{p}' \tilde{\mathcal{V}}_w(\mathbf{r}, \mathbf{p} - \mathbf{p}') \tilde{f}_w(\mathbf{r}, \mathbf{p}') + \Gamma \tilde{f}_w \end{aligned} \quad (26)$$

where the introduction of the exponential factor brings about an additional interaction mechanism, with a constant coupling Γ . The WP's method is then applied to Eq. (24). This algorithm makes possible the inclusion in the simulation of a higher number of scattering events, thus allowing to reach longer and physically more significative simulation times, and, at the same time, introduces a “natural distribution” for the flight duration. Using the transfer function $\tilde{\mathcal{V}}_w$, the time integration and the Neumann expansion, implies that now between one scattering and another the factor $e^{-\Gamma(t_i-t_j)}$ has to be added (that is canceled with the weight if Γ is used to generate the free flights, as in the traditional MC), and at each scattering event a choice has to be performed

(with arbitrary probabilities) whether selecting a physical scattering or a self-scattering. Then the probabilities used are accounted for in the evaluation of the weight.

A MC simulation of the Wigner equation with the above splitting between classical force, whose effect is included in the free flight dynamics, and quantum effects as collision integral and including the quantum self-scattering mechanism, is at present under development. The algorithm implemented, for a case where no phonons are included, and preliminary results are presented in the following sections.

6. A Possible Algorithm

The mathematical model presented in Section 4 shows that the way of selecting the WP's is completely free. As a consequence a number of different algorithms can be devised according to the specific problem to be faced. Here we describe a specific one, among those we have developed, where the separation between the classical force and the quantum corrections is accounted for, the quantum self-scattering is included, and, for simplicity, no phonon-scattering is considered. In particular we describe a backward procedure for the case of electrons in a one-dimensional device, where an applied electric field E and a potential profile $V(x)$ are present.

1. Definition of the data of the physical system, of the simulation time t (of the order of 10^{-11} s) and of $\Gamma = 1/\tau$, to be used for the quantum self-scattering contribution. The device is supposed to be empty at $t = 0$.
2. Definition of the potential profile by means of an analytical expression (steps and barriers can be well described by means of combinations of Fermi functions).
3. Selection of a specific point (x, p) of the phase-space, where the WF has to be evaluated.
4. The weight of each paths is initialized to weight = 1.
5. Starting from the final point (x, p) of the phase-space at time t , a free flight dt is considered (dt is taken a constant and of the order of 10^{-17} s). A new point (x', p') at time $t - dt$ is determined according to classical equation of motion:

$$\begin{cases} x'(t - dt) = x - \frac{p}{m} dt + \frac{1}{2} \frac{(eE + \nabla V(x))}{m} dt^2 \\ p'(t - dt) = p + \frac{(eE + \nabla V(x))}{m} dt \end{cases}. \quad (27)$$

- 6a. If the boundary is reached a value is assigned to the WF, given by

$$f_w(x, p, t) = f^{(c)}(x_c, p_c) \times \text{weight}$$

where $f^{(c)}(x_c, p_c)$ is the assigned boundary condition (x_c being the position of the boundary and p_c the value of the momentum when the boundary is reached). The simulation proceeds from step 10)

- 6b. If the simulation time reaches the initial value $t = 0$, $f_w(x, p, t)$ is set zero, since as initial condition we assume $f^{(c)}(x_c, p_c)$ on the boundary at any time, and 0 inside the device at $t = 0$. The simulation proceeds from step 10).
7. Once the new coordinates of the simulative particle are evaluated, a choice is made about performing or not a scattering according to the scattering probability $P_s = (1 - e^{-dt/\tau})$
8. If no scattering is selected, the weight of the path is multiplied by the factor $\exp(-dt/\tau)/(1 - P_s)$, x' and p' are substituted by x and p and the simulation proceeds from step 5).
9. In case a scattering event is chosen the weight of the path is multiplied by the factor $1/P_s$, and a further selection is performed between scattering with the potential profile (with probability $P_V(x') = \int dp |\tilde{\mathcal{V}}_w(x', p)| \times dt/P_s$, where Γ is chosen in such a way that $P_V < 1$ always) and a self-scattering (with probability $P_{self} = 1 - P_V(x')$).
- 9a. In the self-scattering case the weight of the path is multiplied by the factor $\exp(-dt/\tau)/P_{self}$, x' and p' are substituted by x and p and the simulation proceeds from point 5).
- 9b. If a scattering with the potential is chosen then the exchanged momentum Δp is determined (with a probability $P(\Delta p) \propto |\tilde{\mathcal{V}}_w(x', \Delta p)|$), the weight of the path is multiplied by the factor $\exp(-dt/\tau)/P(\Delta p)/P_V(x')$, and the simulation proceeds from step 5) with phase-space coordinates ($x = x'$, $p = p' - \Delta p$).
10. The procedure is repeated until the established number of paths is reached. Then an average over the number of simulated paths is performed.
11. Back to step 3) a new point of the phase-space is considered.

7. Preliminary Results

In the following we present few preliminary results of the above algorithm. The sample device is a 200 nm wide

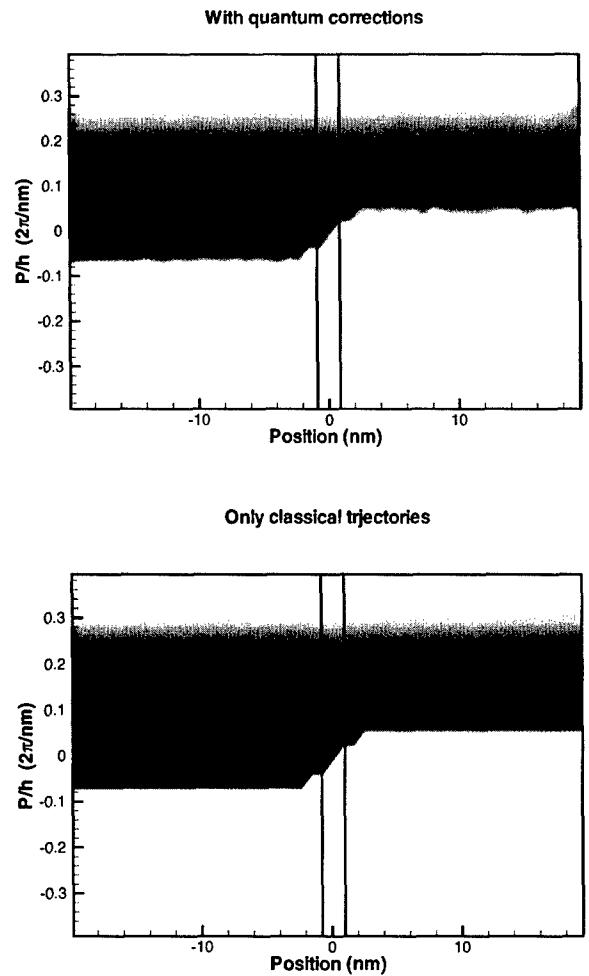


Figure 1. 2D graphs of the WF for the case of a barrier 2 meV high and 2 nm wide. A comparison is performed between the case in which the quantum corrections are included (top) and the case where only classical orbits are considered (bottom). The two vertical lines indicate the position of the potential barrier, while the horizontal one indicates the classical threshold, determined by the barrier height. Darker tones of grey indicate higher values of the WF, while the zero value is white.

system of GaAs, with a potential barrier centered in the middle of the system itself. In the present simulation electrons are entering the device from the left boundary (negative values of x , positive values of p).

Figure 1 shows a 2D graph of the WF for the case of a barrier 2 meV high and 2 nm wide. A comparison is performed between the case in which the quantum corrections are included (top) and the case where only classical orbits are considered (bottom). The steps in the barrier zone are due to the discretization in the momentum space. While in the classical plot the WF is zero in the positive x region below the barrier height

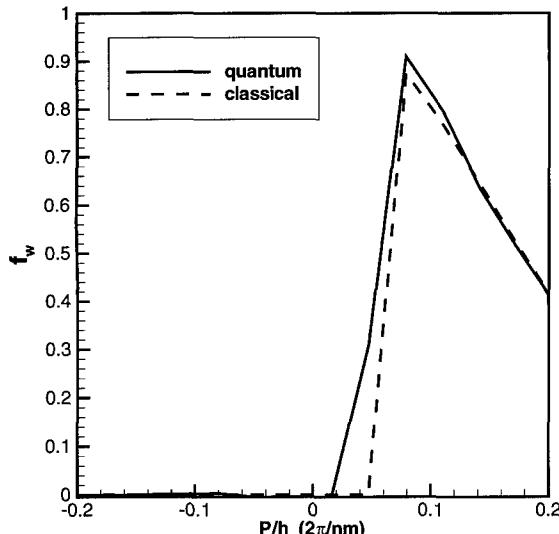


Figure 2. comparison between the classical WF (dashed line) and the WF including quantum corrections (solid line) for a specific value of the position $x = 7$ nm. The parameter of the system are the same as in Fig. 1.

threshold, the quantum result, due to tunneling phenomena, shows positive contribution to the WF in this classically forbidden region.

This effect is more noticeable in Fig. 2, where the comparison between the classical distribution function (dashed line) and the quantum WF (solid line) is presented for a specific value of the position $x = 7$ nm. The parameter of the system are the same as in Fig. 1. Figure 3 shows the same comparison of Fig. 2, for a Barrier of 3 meV.

While from the qualitative point of view these results give us confidence on the correctness of the method, they are still unsatisfactory as practical tool from the quantitative respect. So far problems related to the convergence of the variance did not allow us to study more realistic physical situations. On the other hand, just the above mentioned great flexibility of the WP's approach make us confident that this limitation can be overcome just finding out a more efficient way of selecting the paths. In this direction we are now focusing our research efforts.

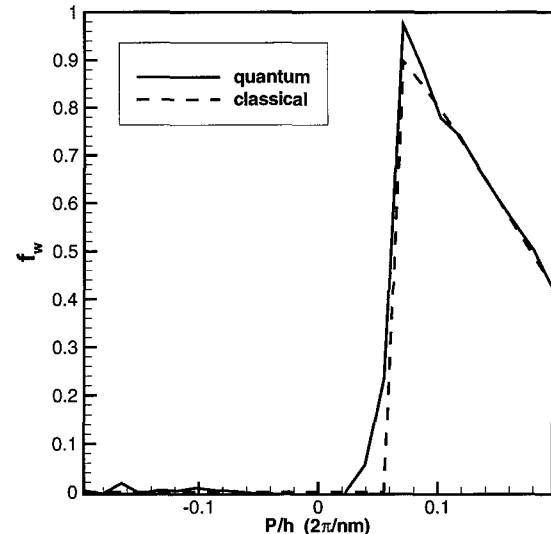


Figure 3. Same as in Fig. 3, for the case of a 2 nm wide, 3 meV high, barrier.

Acknowledgments

This work has been supported by the U.S. Office of Naval Research (contract No. N00014-98-1-0777), by the MIUR, and by the CNR under the project MADESS II.

References

- Bertoni A., Bordone P., Brunetti R., and Jacoboni C. 1999. J. Phys.: Condens. Matter 11: 5999.
- Brunetti R., Jacoboni C., and Rossi F. 1989. Phys. Rev. B 39: 10781.
- Jacoboni C., Brunetti R., Bordone P., and Bertoni A. 2001. In Topics in High Field Transport in Semiconductors, edited by Brennan K. and Ruden P.P. World Scientific Publishing Company, Singapore, p. 25.
- Lozovik Y.E. and Filinov A.V. 1999. JETP 88: 1026.
- Menziani P., Rossi F., and Jacoboni C. 1989. Solid-State Electronics 32: 1807.
- Pascoli M., Bordone P., Brunetti R., and Jacoboni C. 1998. Phys. Rev. B 58: 3503.
- Rossi F. and Jacoboni C. 1992. Europhys. Lett. 18: 169.
- Rossi F., Jacoboni C., and Nedjalkov M. 1994. Semicond. Sci. and Technol. 9: 934.
- Rossi F., Poli P., and Jacoboni C. 1992. Semicond. Sci. Technol. 7: 1017.



Parallelization of the Nanoelectronic Modeling Tool (NEMO 1-D) on a Beowulf Cluster

GERHARD KLIMECK

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

gekco@jpl.nasa.gov

Abstract. NEMO's main task is the computation of current-voltage (I-V) characteristics for resonant tunneling diodes (RTDs). The primary model for high performance RTDs is the full band sp₃s* tight binding simulation, which is based on a numerical double integral of energy and transverse momentum over a transport kernel at each bias point. A full charge self-consistent simulation invoking this model on a single CPU is prohibitively expensive, as the generation of a single I-V curve would take about 1–2 weeks to compute. Simplified charge self-consistent models, eliminating the numerical momentum integral for the quantum mechanical charge self-consistency, followed by a single pass double integration for the current, have been used in the past. However, Computation on a parallel computer now enables the thorough exploration of quantum mechanical transport including charge self-consistency effects within the entire Brillouin zone based on the double integral. Various parallelization schemes (fine, coarse, and mixed) are presented and evaluated in their performance. Finally a comparison to experimental data is given.

Keywords: NEMO, heterostructures, tunneling, parallel, cluster, tight binding, adaptive mesh

1. Introduction

1.1. Nanoelectronic Modeling (NEMO)

The Nanoelectronic Modeling tool¹ (NEMO) was developed as a general-purpose quantum mechanics-based 1-D device design and analysis tool from 1993–98 by Texas Instruments/Raytheon. NEMO enables the fundamentally sound inclusion of the required physics to study electron transport in resonant tunneling diodes (RTDs): bandstructure, scattering, and charge self-consistency based on the non-equilibrium Green function approach. The theory used in NEMO and the major simulation results are published (see Klimeck *et al.* 1997, Bowen *et al.* 1997 and references therein).

NEMO's main task is the computation of current-voltage characteristics for high performance resonance tunneling diodes at room temperature. The primary transport model used for these simulations is based on a sp₃s* tight binding representation of the non-parabolic

bands and the integration of a momentum and energy dependent transport kernel. The total energy integral and the transverse momentum integral extends over the occupied states in the RTD. The energy integral typically covers about 1 eV, and the transverse momentum typically extends to about 10% of the Brillouin zone from the Γ point for typical InGaAs/InAlAs RTDs on an InP substrate. The physical model has been discussed in detail (Bowen *et al.* 1997) before. Previous simulations (Klimeck *et al.* 1997, Bowen *et al.* 1997) which agreed quantitatively with experiment were lacking one major feature: the models in which the current and the potential/charge were calculated were not self-consistent with each other. The parallelization of NEMO described and characterized in this paper enables such self-consistent simulations.

1.2. Parallelization on Cluster Computers

The availability of relatively cheap PC-based Beowulf clusters offers research and/or development groups an affordable entry of into massively parallel computing.

To whom correspondence should be addressed.

Our research group at JPL has developed, implemented, and maintained various generations of clusters (Cwik *et al.* 2001). The benchmarks that are presented in this paper were run on a 32 node, 64 CPU Pentium III 933 MHz cluster connected on a standard 100 Mbps network. Parallel code was developed using the Message Passing Interface (MPI).

2. Code Parallelization

2.1. The Transport Kernel

NEMO's core numerical task is the integration of a transport kernel, \mathcal{K} , at the n th bias voltage to obtain current, I_n , and charge $N_{n,i}$ on every site i . That kernel is dependent on the total energy, E , the transverse momentum, k , the potential profile and applied voltage, $V_{n,i}$, and the charge at the previous bias voltage, q_{n-1} .

$$\{I_n, N_{n,i}\} = \int dE \int kdk \mathcal{K}(E, k, V_{n,i}, q_{n-1}) \quad (1)$$

$$\{I_n, N_{n,i}\} \approx \int dE \mathcal{K}'(E, V_{n,i}) \quad \text{Tsu-Esaki} \quad (2)$$

Equation (2) stems from the typical Tsu-Esaki assumption (Tsu and Esaki 1973) of parabolic transverse subbands which enables an analytic integration over k . Equations (1) and (2) result in significantly different currents (Bowen *et al.* 1997) and charge distributions. The charge N_i must be computed self-consistently with the electrostatic potential through Poisson's equation. Different charge distributions, N_i , will result in different potential distributions, $V_{n,i}$, which will in turn result in different current distributions, I_n . However, the best (Klimeck *et al.* 1997, Bowen *et al.* 1997) that was done due to realistic time constraints so far was to compute N_i from Eq. (2) self-consistently with $V_{n,i}$, and then perform a one-pass calculation with a fixed potential to obtain a current using Eq. (1). Parallelization of NEMO makes a fully charge self-consistent simulation possible by moving the computation time down to 10–20 hours on an adequately sized cluster (16–32 nodes).

The benchmark I-Vs presented in this paper are based on a semi-classical charge-self-consistent potential (Thomas-Fermi) with 70 bias points, including 21 momentum points resolving up to 7% of the Brillouin zone around Γ . The integral over total energy is performed with an adaptive search algorithm (Klimeck *et al.* 1998) that starts from 200 energy nodes and

resolves resonances in the transmission and the charge density through iterative refinement.

2.2. Parallelization Around Bias Points

Typical I-Vs span a voltage of 0.7 V, which results with a typical resolution of 0.01 V in 70 bias points. If the simulation does not need to include any charge accumulation effects from one bias point to the next (hysteresis or switching), then the dependence on q_{n-1} in Eq. (1) can be neglected and all bias points n can be considered independent of each other. This simplification suggests a parallelization scheme where the individual bias points are farmed out to different CPUs. This scheme implies minimal communication between the CPUs and minimal interference of the algorithm with the remaining 250,000 lines of C, FORTRAN, and F90 code in NEMO.

Various implementations for such an outer loop parallelization are possible. In the simplest case, all the bias points are distributed to N CPUs in a single communication step and the results could be gathered in a second communication step. Such a scheme may be hampered by a load balancing problem, since the computation time needed for each bias point may vary from one to the next for various reasons: the energy range in which transport is computed is bias dependent, the charge self-consistency may require a different number of iteration steps at different biases (especially at the I-V turn-off), and in a cluster of workstations the CPU speed may vary. To treat this load balancing problem and to minimize the communication contention with the central CPU, a master/slave approach was chosen, where the master's job is to distribute single bias points to available slaves and to gather completed I-V points from slaves. Such an approach can be very inefficient on a few CPUs, since the master is mostly sitting idle, waiting for results to be returned. However, MPI can be instructed such that a master and a slave run on a single CPU simultaneously, where the master CPU only gets real CPU time when it is needed for communication. In the benchmarks performed here a master was assigned to its own CPU.

The line marked with circles in Fig. 1(a) indicates the actual CPU times that were obtained on our cluster as a function of number of processors. Almost perfect scaling with processors up to 15 CPUs can be observed, when a step-like structure becomes apparent for an increasing number of CPUs. At 24 and 36 CPUs almost perfect scaling can be seen, which can be explained

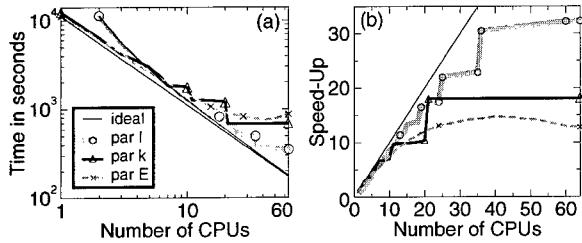


Figure 1. (a) Total time for the computation of an I-V without charge self-consistency (only semi-classical charge self-consistency) as a function of number of CPUs used in the parallel algorithms. Ideal performance is depicted as a straight line on a log-log scale. 70 bias points (I), 21 k points, adaptive E grid. Parallelization in I, k , and E. (b) Speed-up due to parallelization compared to the single CPU performance.

with the finite number of 70 bias points that are computed. To illustrate this point more clearly, Fig. 1(b) shows the speed-up due to parallelization as a function of number of CPUs. From 24 to 35 CPUs, at least one CPU must compute 3 bias points, and, although some CPUs finish earlier after computing just 2 bias points, the whole I-V is not finished until all CPUs report their results. Similar load imbalance with 1 or 2 bias points per CPU causes the step from 36 to 64 CPUs. *If the number of bias points is increased to several hundred, almost perfect scaling without the steps in Fig. 1 is observed (not shown here).* However, a realistic number of bias points was chosen to show the problems with the parallelism.

Figure 1 shows good efficiency in the parallelism over bias points. However, from a device research point of view, it is often very instructive to study a single bias point in detail, and it is desirable to get results as fast as possible. Additionally, in calculations that consider charge accumulation, the dependence on q_{n-1} in Eq. (1) can not be neglected, the bias points are therefore not independent of each other and parallelization around the bias points may result in an incorrect I-V. A parallelization that is finer grain than parallel voltage points is therefore desirable.

2.3. Parallelization of Transverse Momentum Integral

The integral over the total energy, E , of Eq. (1) results in an integrand, $J(k)$, that is still a function of transverse momentum, k . This integrand can be shown (Klimeck, Bowen and Boykin 2001a, 2001b) to be typically monotonically decreasing from $k = 0$. Only in

rare cases is the electron transport anisotropic (Klimeck 2001), implying that the function $J(k)$ can be resolved well with only a few numerical nodes, typically 15–29. The benchmark simulations are based on 21 points.

Since the workload for each k point is about the same, a simple parallelization scheme was chosen: the k points are distributed to all available CPUs. With only 21 k points available, good scalability of the parallel algorithm is limited to 21 CPUs, with a strongly visible load imbalance step at 11 CPUs (triangles in Fig. 1).

The parallelization around k points does not appear to be very advantageous in the benchmark shown here, except for the commensurability points at 11 and 21 CPUs. Note, however, that simulations of hole transport (Klimeck, Bowen and Boykin 2001a, 2001b) required about 150 k points due to the large anisotropy in $J(k)$, and the parallelization around k points was essential to obtain results at a single bias point.

2.4. Parallelization of Total Energy Integral

The integral of the transport kernel over total energy is the lowest level integral that is evaluated in NEMO. For high performance RTDs, where the resonances are not narrow in energy, this integral is typically² performed in an adaptive Simpson-type 3 and 5 point algorithm, where 2 energy points are added to the 3 point integral to evaluate the change of the overall integral value. The work-load is identical for each energy suggesting a complete distribution of all new required refinement energies to the available CPUs in one communication step. In a typical structure, only one or two resonances must be resolved well within the energy range of interest. The final refinement steps will therefore request two or four new energies to be computed. The limited number of new energy nodes requested towards the end of the refinement limits the performance of this energy parallelism. Figure 1 (crosses) shows a respectable scaling with increasing number of CPUs up to 20 CPUs. Increased communication costs for large numbers of CPUs actually degrades the performance beyond 40 CPUs on this cluster with a slow 100 Mbps network. Preliminary results on our new cluster, which is equipped with a 2 Gbps network, show significantly improved scaling of this fine grain parallelism.

2.5. Multiple Levels of Parallelism

The coarse and the medium grain (I and k) parallel schemes show significant load balancing problems for

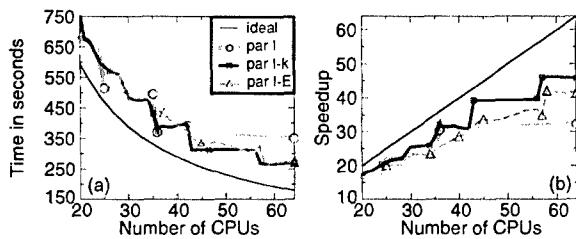


Figure 2. (a) Total compute time as a function of number of CPUs for three different parallelization schemes. Simultaneous parallelization in I-k and I-E improves performance over simple parallelization in I. (b) Speedup due to parallelization measured against single CPU performance.

larger numbers of CPUs in a realistic I-V computation. The fine grain parallelism (*E*) is communication limited and load-balancing limited. To enable a speed-up of a realistic I-V calculation a combination of these parallel algorithms has been implemented. Each bias point (*I*) can now be assigned to a group of CPUs, this *I* group can be subdivided into different groups of momentum points (*k*), and these *k* groups can be subdivided into groups of energy points (*E*). Four parallel schemes are therefore possible: I-k, I-E, k-E, and I-k-E. The user can specify the desired level of parallelism and the size of the groupings. An automated assignment of group sizes tries to select large parallel groups starting from the coarse level parallelism. Figure 2 compares the performance of parallelism in I-k and I-E to the parallelism in I. At 64 CPUs a significant improvement of the speed-up from 32 to 45 is achieved. Some commensurability steps in the performance as a function of number of CPUs are still visible suggesting the possibility of improvement on the automated CPU grouping algorithm.

3. Comparison to Experiment

The structure considered here is part of the NEMO InP testmatrix (Klimeck *et al.* 1997). The sample consists of an undoped central structure InGaAs/InAlAs/InGaAs/InAlAs/InGaAs with 7/17/17/17/7 monolayers, respectively. The central structure is surrounded by 50 nm low doping (10^{18} cm^{-3}) buffer and high doping ($5 \times 10^{18} \text{ cm}^{-3}$) contacts.

The simulations in the benchmark presented in Figs. 1 and 2 are based on a Thomas-Fermi (TF) semi-classical charge self-consistent potential. The resulting I-V curve is compared in Fig. 3(a) in dashed line to experimental data (thick solid line). To achieve better agreement on the overall peak shape, a simulation

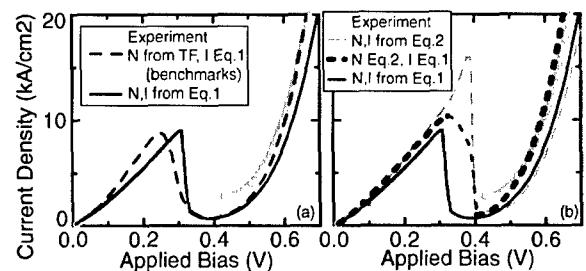


Figure 3. Computed I-V characteristics for a InGaAs/InAlAs RTD compared to experimental data. (a) Benchmark simulations using semi-classical (Thomas-Fermi) self-consistency and full quantum (Hartree) self-consistency. Hartree self-consistency represents the shape of I-V properly. (b) Improved simulation capabilities: (1) charge and current from Eq. (2), (2) charge from Eq. (3), current from Eq. (1), and (3) charge and current from Eq. (1).

must include (Klimeck *et al.* 1997, Bowen *et al.* 1997) quantum charge self-consistently in the potential calculation. Such a fully self-consistent simulation using the 10 band sp_{3s}* tight binding model is shown here for the first time in Fig. 3(a) with a thin solid line.

A simulation solely based on Eq. (2) shown in Fig. 3(b) with a thin dashed line shows a significant current overshoot (Bowen *et al.* 1997) at the I-V turn-off. A single pass computation of the current with Eq. (1) using the self-consistent potential of Eq. (2) results in a smoothing (Bowen *et al.* 1997) of the current spike. The unphysical rounding in the NDR (thin dashed line) was neglected in previous runs (Klimeck *et al.* 1997, Bowen *et al.* 1997). With the new parallel NEMO code, the current and the charge can now be computed fully self-consistently (thin solid line).

4. Summary

This work shows the utility of low-cost, high performance Beowulf clusters for the design and characterization of electronic devices using physics-based simulation software. Various parallelization schemes (coarse, medium, fine, and mixed grain) are shown for the NEMO 1-D simulator resulting in the capability to simulate for the first time full charge self-consistent simulations including full bandstructure effects within a significant portion of the Brillouin zone using the sp_{3s}* tight binding model.

Acknowledgments

The work described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of

Technology under a contract with the National Aeronautics and Space Administration. The supercomputer used in this investigation was provided by funding from the NASA Offices of Earth Science, Aeronautics, and Space Science. I would also like to acknowledge fruitful collaborations that lead up to this work within the NEMO team consisting of Dr. R. Chris Bowen, Dr. Roger Lake and Dr. Timothy B. Boykin. I would also like to thank Dr. Charles Norton and Dr. Victor Decyk for their help with MPI and other parallel coding issues, as well as Frank Villegas and T. Wack for the review of the manuscript.

Notes

1. See <http://hpc.jpl.nasa.gov/PEP/gekco/nemo> or search for NEMO on <http://www.raytheon.com>.
2. In structures where the barriers are thick, such as quantum wells, or hole structures (Klimeck, Bowen and Boykin 2001a), the resonances are very sharp in energy ($<0.01 \mu\text{eV}$) and the adaptive

algorithm can be shown to be quite ineffective compared to a resonance finding-based algorithm (Klimeck *et al.* 1998).

References

- Bowen R.C., Klimeck G., Lake R.K., Frenley W.R., and Moise T. 1997. J. Appl. Phys. 81: 3207.
 Cwik T., Klimeck G., McAuley M., Norton C., Sterling T., Villegas F., and Wang P. 2001. The use of clustercomputer systems for NASA/JPL Applications. AIAA Space 2001 Conference and Exposition, Albuquerque, New Mexico.
 Klimeck G. 2001. Physica Status Solidi 226: 9.
 Klimeck G., Bowen R.C., and Boykin T.B. 2001a. Superlattices and Microstructures 29: 187.
 Klimeck G., Bowen R.C., and Boykin T.B. 2001b. Phys. Rev. B 63: 195310.
 Klimeck G., Boykin T.B., Bowen R.C., Lake R., Blanks D., Moise T.S., Kao Y.C., and Frenley W.R. 1997. In the 1997 55th Annual Device Research Conference Digest, (IEEE, NJ, 1997), p. 92.
 Klimeck G., Lake R., Bowen R.C., Fernando C.L., and Frenley W.R. 1998. VLSI Design 6: 107.
 Tsu R. and Esaki L. 1973. Appl. Phys. Lett. 22: 562.



Towards Fully Quantum Mechanical 3D Device Simulations

M. SABATHIL, S. HACKENBUCHNER, J.A. MAJEWSKI, G. ZANDLER AND P. VOGL

*Walter Schottky Institute and Physics Department, Technical University of Munich, Am Coulombwall,
85748 Garching, Germany*

Abstract. We present a simulator for calculating, in a consistent manner, the realistic electronic structure of three-dimensional heterostructure quantum devices under bias and its current density close to equilibrium. The electronic structure is calculated fully quantum mechanically, whereas the current is determined by employing a semiclassical concept of local Fermi levels that are calculated self-consistently. We discuss the numerical techniques employed and present illustrative examples that are compared with quantum transport calculations. In addition, the simulator has been used successfully to study shape-dependent charge localization effects in self-assembled GaAs/InGaAs quantum dots.

Keywords: nano-devices, simulations, quantum dots

1. Introduction

State-of-the-art simulators for semiconductor nanostructures and optoelectronic nano-devices roughly fall into two classes (Anonymous 2000): some models focus on the *equilibrium* electronic structure. They attempt to predict, as accurately as possible, the free and bound charge density as well as optical properties of quantum wells, wires, and dots on a length scale that ranges from nm to μm . Several models of this kind have been developed in the last few years that can deal with fully three-dimensional device geometries, and invoke one-band (Kumar, Laux and Stern 1990, Laux 1987), or several-band $k \cdot p$ models (Grundmann, Stier and Bimberg 1999, Pryor 1998, 1999, Cusack, Briddon and Jaros 1996), tight binding methods (Carlo 2001). The second class of models focus on *current-voltage characteristics* and attempts to solve quantum transport equations, using non-equilibrium Green functions (Lake *et al.* 1997, Klimeck, Lake and Blanks 1998), Wigner functions (Bordone *et al.* 1999, Grubin, Buggeln and Kreskovsky 2000, Ferry, Akis and Vasileska 2000), or the Pauli master equation (Fischetti 1998, 1999). Presently, they are still limited to one spatial dimension and/or put less emphasis on details of the electronic structure. Thus, a simultaneous realistic treatment of the electronic structure and the

quantum transport problem for 3-D structures still poses a challenging task.

In this paper we discuss a simulator that we are currently developing for a wide class of 3-D Si and III-V nano-structures (Hackenbuchner *et al.* 2001). It attempts to bridge the two types of approaches described above, albeit with a stringent limitation that makes it feasible to simulate three-dimensional structures: we solve the electronic structure problem accurately but restrict the current evaluation to situations close to equilibrium where the concept of local quasi-Fermi levels is still justifiable. This approach may be viewed as a low-field approximation to the Pauli master equation (Jones and March 1973). In this paper, we present the numerical methods involved, illustrate this method for simple 1-D situations that allow a detailed comparison with the full Pauli master equation approach that has recently been developed by Fischetti (1998, 1999), and briefly show that this simulator successfully predicts the electronic and optical properties of self-assembled GaAs/InAs quantum dots.

2. Method

The nano-device simulator that we have developed so far solves the 8-band- $k \cdot p$ -Schrödinger-Poisson equation for arbitrarily shaped 3-D heterostructure device

geometries, and for any (III-V and Si/Ge) combination of materials and alloys. It includes band offsets of the minimal and higher band edges, absolute deformation potentials (Van de Walle 1989), local density exchange and correlations (i.e. the Kohn-Sham equations), total elastic strain energy (Pryor *et al.* 1998, Grundmann, Stier and Bimberg 1995) that is minimized for the whole device, the long-range Hartree potential induced by charged impurity distributions, voltage induced charge redistributions, piezo- and pyroelectric charges, as well as surface charges, in a fully self-consistent manner. The charge density is calculated for a given applied voltage by assuming the carriers to be in a *local* equilibrium that is characterized by energy-band dependent local quasi-Fermi levels $E_{Fc}(\mathbf{x})$ for charge carriers of type c (i.e. in the simplest case, one for holes and one for electrons),

$$n_c(\mathbf{x}) = \sum_i |\Psi_{ic}(\mathbf{x})|^2 f\left(\frac{E_{Fc}(\mathbf{x}) - E_{ic}}{k_B T}\right). \quad (1)$$

These local quasi-Fermi levels are determined by global current conservation $\nabla \cdot \mathbf{j}_c = 0$, where the current is assumed to be given by the semi-classical relation $\mathbf{j}_c(\mathbf{x}) = \mu_c n_c(\mathbf{x}) \nabla E_{Fc}(\mathbf{x})$ (see, e.g. (Selberherr 1984)). Recombination and generation processes are included additionally. The carrier wave functions Ψ_{ic} and energies E_{ic} are calculated by solving the multi-band Schrödinger-Poisson equation. The open system is mimicked by using mixed Dirichlet and von Neumann boundary conditions (Fischetti 1998, Lent and Kirkner 1990, Frenzley 1992) at Ohmic contacts.

3. Computational Schemes

For a given nano-structure, the computations start by globally minimizing the total elastic energy (Pryor *et al.* 1998, Grundmann, Stier and Bimberg 1995) using a conjugate gradient method. This determines the piezo-induced charge distributions, the deformation potentials and band offsets. Subsequently, the 8-band-Schrödinger, Poisson, and current continuity equations are solved iteratively. All equations are discretized according to the finite difference method invoking the box integration scheme (Kumar, Laux and Stern 1990, Laux 1987, Selberherr 1984). The irregular rectilinear mesh is kept fixed during the calculations. As a preparatory step, the built-in potential is calculated for zero applied bias by solving the Schrödinger and Poisson equation self-consistently employing a predictor-corrector

approach (Trellakis *et al.* 1997) and setting to zero the electric field at the Ohmic contacts. For applied bias, the Fermi level and the potential at the contacts are then shifted according to the applied potential which fixes the boundary conditions. The main iteration scheme itself consists of two parts. In the first part, the wave functions and potential are kept fixed and the quasi-Fermi levels are calculated self-consistently from the current continuity equations, employing a conjugate gradient method and a simple relaxation scheme. In the second part, the quasi-Fermi levels are kept constant, and the density and the potential are calculated self-consistently from the Schrödinger and Poisson equation. The discrete 8-band-Schrödinger equation represents a huge sparse matrix (typically of dimension 10^5 for 3D-structures) and is diagonalized using the Jacobi-Davidson method (Bai *et al.* 2000) that yields the required inner eigenvalues and eigenfunctions close to the energy gap. We very slightly shift the spin-up and spin-down diagonal Hamiltonian matrix elements with respect to each other in order to avoid degeneracies and guarantee orthogonal eigenstates automatically. To reduce the number of necessary diagonalizations, we employ an efficient predictor-corrector approach (Trellakis *et al.* 1997) to calculate the potential from the nonlinear Poisson equation. In this approach, the wave functions are kept fixed within one iteration and the density is calculated perturbatively from the wave functions of the previous iteration (Trellakis *et al.* 1997). The nonlinear Poisson equation is solved using a modified Newton method, employing a conjugate gradient method and line minimizations. The code is written in Fortran 90 and consists of some 170000 lines by now.

4. One-Dimensional Examples and Comparison to Pauli Master Approach

In this section we illustrate the present method by studying simple one-dimensional examples that we can compare to full Pauli master equation results (Fischetti 1998). As discussed above, our method amounts to calculating the electronic structure of a device fully quantum mechanically, yet employing a semiclassical scheme for the evaluation of the current. As we shall see, the results are close to those obtained by the full Pauli master equation (Fischetti 1998) provided we limit ourselves to situations not too far from equilibrium.

As a first example, we consider a one-dimensional (Si-based) n-i-n structure (Fig. 1) at room temperature

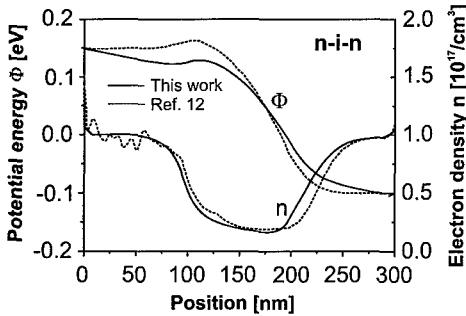


Figure 1. Calculated potential energy and electron density of n-i-n structure, as a function of position inside the structure. Results from presently developed method (full lines) are compared to results from Pauli master equation (Fischetti 1998) (dashed lines).

with 100 nm long *n*-typed regions and a doping concentration of $n = 10^{17} \text{ cm}^{-3}$. The applied bias is 0.25 V. As one may expect, true quantum mechanical effects play little role in this case and both the present, the semiclassical drift-diffusion as well as the Pauli master equation (PME) approach of Fischetti (1998) yield practically identical results for the density and potential profile. We would like to point out, that this good agreement is a nontrivial finding, as we calculate the density quantum mechanically according to Eq. (1) with self-consistently computed local Fermi levels rather than semiclassically. Our current density of $3.6 \times 10^4 \text{ A/cm}^2$ compares well with the PME result of $6.8 \times 10^4 \text{ A/cm}^2$. However, we note that the current is directly proportional to the mobility in our model; changing the mobility (that we treat as constant in this example) therefore changes the value of the current but does not affect the electron density or potential profile.

In Fig. 2 we show results for a Si-SiO₂-Si structure with a 3 nm barrier of 3 eV height and a forward bias of 1 V. Again, our calculated density and potential profile can be seen to agree very well with the PME method (Fischetti 1998). Due to the high barrier, the current density predicted by the PME is only of the order of 10^{-8} A/cm^2 . In our approach, the tunneling current density is still small but significantly higher (10^{-3} A/cm^2). This originates in our simplified treatment of boundary conditions which mimic an open system by linearly combining stationary (non-current carrying) solutions of the Schrödinger equation with von Neumann and Dirichlet boundary conditions at the contacts (Lent and Kirkner 1990, Frensel 1992). For a high barrier near the middle of the device, the minimum electron density for zero or small bias lies near the center of the barrier irrespective of the boundary

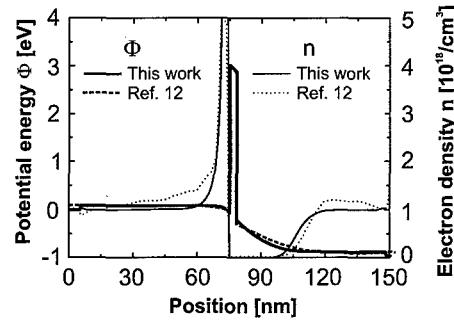


Figure 2. Calculated potential energy Φ and electron density n of Si-SiO₂-Si structure with 3 nm barrier, as a function of position inside the structure. Results from presently developed method (full lines) are compared to results from Pauli-master equation (Fischetti 1998) (dashed lines).

conditions used. This is in contrast to current carrying scattering solutions that travel from left to right (right to left) and reach their minimum at the right (left) side of the barrier. While this difference has practically no effect on the potential and density outside the barrier, it raises the tunneling current significantly for small bias and very high barriers. We are currently generalizing our approach by filtering out scattering states from the stationary Schrödinger solutions and deriving appropriate open boundary conditions, in analogy to very recent work of Laux and Fischetti (unpublished).

Finally, we study the same n-i-n structure as in Fig. 1, but with an additional well of 10 nm width and 0.1 eV depth. Figure 3 shows the potential profile. The well supports three strongly localized states. Quantum mechanics yields a smaller electron density

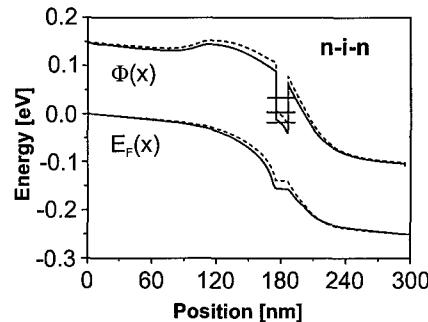


Figure 3. Calculated potential energy Φ and local electron Fermi level E_F with a 10 nm wide quantum well for a n-i-n structure as a function of position. The full line is obtained with the present method, whereas the dashed line shows the classical drift-diffusion solution. Also shown are the positions of the three localized energy levels of the well.

in the well region than the classical Thomas-Fermi (drift-diffusion) solution which is a physically plausible result. This leads to a larger potential drop across the well region which in turn results in a higher current density J . The classical drift-diffusion solution gives $J = 2.2 \times 10^4 \text{ A/cm}^2$ whereas the present model gives $J = 2.5 \times 10^4 \text{ A/cm}^2$. The localized states themselves contribute very little to the current. As a consequence, the local Fermi level is practically constant across the bound states which provides an *a posteriori* justification of invoking non-local quantum mechanics concurrently with the semi-classical concept of local Fermi levels. In fact, no conflict arises for situations close to equilibrium since the spatial variation of the occupancy of any given eigenstate turns out to be negligible for three reasons: (i) Deeply bound states do not contribute to the current and thus do not lead to a gradient of the Fermi level; (ii) the Fermi level has the largest variation in regions where the density is very low (within barriers, for example); (iii) very extended states that are treated formally as bound states in our method are either not occupied because of their high energy, or occur in regions of high density (near contacts, for example) where the quasi-Fermi level is nearly constant.

In summary, the electronic density and potential calculated with the presently developed scheme as a function of applied bias agrees excellently with the more rigorous Pauli master equation approach. More work will be needed, on the other hand, to model ultrasmall tunneling currents accurately.

5. Piezoelectric Fields and Electron-Hole Localization in Quantum Dots

We have applied our simulator to study theoretically single-quantum-dot photodiodes consisting of self-assembled InGaAs quantum dots with a diameter of 30–40 nm and heights of 4–8 nm that are embedded in the intrinsic region of a Schottky diode (Hackenbuchner *et al.* 2001). Recent experiments (Findeis *et al.* 2001, Fry *et al.* 2000) have indicated inverted electron-hole alignments with the electron at the base and the hole at the top of the dot, in contrast to what earlier theoretical calculations (Pryor 1998, 1999, Cusack, Briddon and Jaros 1996, Jones and March 1973) had predicted. Our study reveals that the elastic strain and accompanying piezoelectric fields strongly depend on the geometric shape and alloy composition of dots. We find, in agreement with (Grundmann, Stier

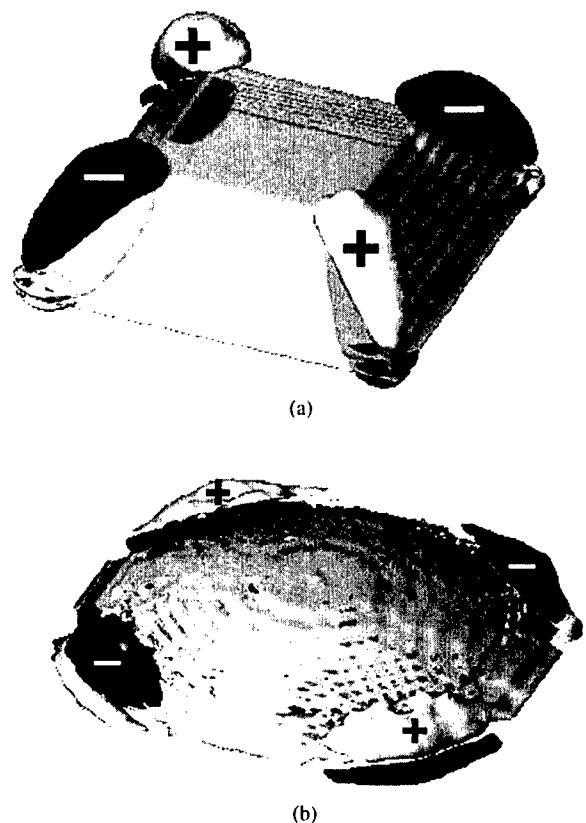


Figure 4. Side view of calculated (a) pyramid-shaped and (b) lens-shaped InGaAs quantum dots. The height in the center is 8 nm for both dot shapes. The light and dark regions are iso-surfaces of the positive and negative piezoelectric charge densities, respectively, with a magnitude of $5 \times 10^{18} \text{ e/cm}^3$.

and Bimberg 1999), dots of pyramidal shape to possess large piezoelectric polarization charges in the corners that lead to a strong hole confinement near the dot edges (Fig. 4(a)). By contrast, lens-shaped dots of similar size have much weaker charges and are found to lead to electron and hole states near the center of the lens which significantly improves the exciton absorption (Fig. 1(b)). Importantly, as we have shown elsewhere (Hackenbuchner *et al.* 2001), we find the hole to be localized at the apex (top) and the electron at the base exactly as has been suggested by Fry *et al.* (2000).

Acknowledgments

Financial support by the Deutsche Forschungsgemeinschaft and by the Office of Naval Research under Contract No. N00014-01-1-0242 is gratefully

acknowledged. We are indebted to Frank Stern for helpful discussions in the early stage of the project.

References

- Anonymous. 2000. In: Proc. of the 7th Int. Workshop on Computational Electronics, Glasgow, Scotland, May 22–25, 2000 and papers therein.
- Bai Z., Demmel J., Dongarra J., Ruhe A., and van der Vorst H. (Eds.). 2000. Templates for the Solution of Algebraic Eigenvalue Problems. SIAM, Philadelphia.
- Bordone P., Pascoli M., Brunetti R., Bertoni A., Jacoboni C., and Abramo A. 1999. Phys. Rev. B 59: 3060.
- Cusack M.A., Briddon P.R., and Jaros M. 1996. Phys. Rev. B 54: R2300.
- Di Carlo A. 2001. Proc. 12th Int. Conf. on Nonequilibrium Carrier Dynamics in Semiconductors, Santa Fe, 2001, also to appear in Physica B.
- Ferry D.K., Akis R., and Vasileska D. 2000. Electron Device Meeting, IEDM 2000 Tech. Digest Papers, pp. 287–290.
- Findeis F., Baier M., Beham E., Zrenner A., and Abstreiter G. 2001. Appl. Phys. Lett. 78: 2958.
- Fischetti M.V. 1998. J. Appl. Phys. 83: 270.
- Fischetti M.V. 1999. Phys. Rev. B 59: 4901.
- Frensley W.R. 1992. Superlattices and Microstructures 11: 3470.
- Fry P.W., Itskevich I.E., Mowbray D.J. *et al.* 2000. Phys. Rev. Lett. 84: 733.
- Grubin H.L., Buggeln R.C., and Kreskovsky J.P. 2000. Superlattices and Microstructures 27: 533.
- Grundmann M., Stier O., and Bimberg D. 1995. Phys. Rev. B 52: 11969.
- Grundmann M., Stier O., and Bimberg D. 1999. Phys. Rev. B 59: 5688.
- Hackenbuchner S., Sabathil M., Majewski J.A., Zandler G., Vogl P., Beham E., Zrenner A., and Lugli P. 2001. Proc. 12th Int. Conf. on Nonequilibrium Carrier Dynamics in Semiconductors, Santa Fe, 2001, also appear Physica B.
- Jones W. and March N.H. 1973. Theoretical Solid State Physics. Wiley, Vol. 2.
- Klimeck G., Lake R., and Blanks D.K. 1998. Phys. Rev. B 58: 7279.
- Kumar A., Laux S.E., and Stern F. 1990. Phys. Rev. B 42: 5166; Laux S.E. 1987. In: Miller J.J.H. (Ed.), Proc. 5th Int. Conf. on Numerical Analysis of Semiconductor Devices and Integrated Circuits (NASECODE V), Boole, Dun Laoghaire, Ireland, pp. 270–275.
- Lake R., Klimeck G., Bowen R.C., and Jovanovic D. 1997. J. Appl. Phys. 81: 7845; Project “NEMO”, <http://www.cfdrc.com/nemo/index.html>.
- Laux S. and Fischetti M.V. Presented at this conference, unpublished.
- Lent C.S. and Kirkner D.J. 1990. J. Appl. Phys. 67: 6353.
- Pryor C. 1999. Phys. Rev. B 57: 7190.
- Pryor C. 1999. Phys. Rev. B 60: 2869.
- Pryor C., Kim J., Wang L.W., Williamson A.J., and Zunger A. 1998. J. Appl. Phys. 83: 2548.
- Selberherr S. 1984. Analysis and Simulation of Semiconductor Devices. Wien, Springer.
- Trellakis A., Galick A.T., Pacelli A., and Ravaioli U. 1997. J. Appl. Phys. 81: 7880.
- Van de Walle C. 1989. Phys. Rev. B 39: 1871.



Simulation of Field Coupled Computing Architectures Based on Magnetic Dot Arrays

GYÖRGY CSABA AND WOLFGANG POROD

Center for Nano Science and Technology, University of Notre Dame, Notre Dame, IN 46556, USA

Abstract. In this paper, we demonstrate that field-coupled nanomagnets can be used for digital information processing. The operation of logic devices is based on a QCA-like architecture, where information propagates by magnetostatic interaction between individual magnetic dots. Micromagnetic simulations indicate that simple logic gates function properly. Efficient design tools, based on the single-domain approximation are developed.

Keywords: magnetic nanocomputing, QCA, micromagnetic design, SPICE

1. Introduction

Magnetic data storage research is currently exploring the possibilities of storing a single bit of information in a single-domain ferromagnetic particle. Magnetic interaction between these particles might cause the loss of information which, of course, is an undesirable characteristic in a storage device. In this paper, we will show, how one can take advantage of such interactions and how to utilize them for information processing.

The architecture is based on the idea of Field-Coupled Computing, i.e. using electric or magnetic interaction between nanosystems to perform computational tasks. This is similar to the *Quantum-Dot Cellular Automata* (QCA) concept, which was originally proposed for Coulomb-coupled quantum dots (Lent *et al.* 1993). Several nanosystems have been studied as building blocks of QCA: Single-Electron QCA's have already been demonstrated (Snider *et al.* 1999), molecular structures have been proposed, and the feasibility of magnetic QCA was experimentally verified (Cowburn and Welland 2000).

Here, we give the first overview of the design of nanomagnetic logic devices. We will introduce the ‘adiabatic control’ of magnetic nanostructures, and propose that appropriate external field control can reliably put the nanomagnet system into the desired state. We will present micromagnetic simulations of functioning

devices, and use the single-domain approximation as a design tool.

2. The Classical Theory of Micromagnetics

This well-known theory exhaustively describes the behavior of ferromagnetic materials, if the size and time scale of interest is large enough that quantum-mechanical effects (i.e. the exchange interaction) can be treated quasi-classically. This quasi-classical approach works for spatial dimensions larger than few nanometers, and times longer than picoseconds. The Landau-Lifshitz equation describes, how the magnetization $\mathbf{M}(\mathbf{r}, t)$ changes under the influence of an effective field $\mathbf{H}_{\text{eff}}(\mathbf{r}, t)$:

$$\frac{\partial \mathbf{M}(\mathbf{r}, t)}{\partial t} = -\gamma \mathbf{M}(\mathbf{r}, t) \times \mathbf{H}_{\text{eff}}(\mathbf{r}, t) - \frac{\alpha \gamma}{M_s} [\mathbf{M}(\mathbf{r}, t) \times \mathbf{M}(\mathbf{r}, t) \times \mathbf{H}_{\text{eff}}(\mathbf{r}, t)] \quad (1)$$

Where γ is the gyromagnetic ratio, α is the damping constant and M_s is the saturation magnetization of the material (Hubert and Schafer 1998). The effective field consists of four parts:

$$\begin{aligned} \mathbf{H}_{\text{eff}}(\mathbf{r}, t) = & \mathbf{H}_{\text{dip}}(\mathbf{r}, t) + \mathbf{H}_{\text{ext}}(\mathbf{r}, t) + \mathbf{H}_{\text{exch}}(\mathbf{r}, t) \\ & + \mathbf{H}_{\text{anis}}(\mathbf{r}, t) \end{aligned} \quad (2)$$

$\mathbf{H}_{\text{dip}}(\mathbf{r}, t)$ represents the dipole-dipole interactions and $\mathbf{H}_{\text{ext}}(\mathbf{r}, t)$ denotes the external magnetic field. Both are ‘real’ magnetic fields, which result from the solution of Maxwell’s equations. $\mathbf{H}_{\text{exch}}(\mathbf{r}, t)$ models the exchange interactions between the moments, and is proportional to the Laplacian of the magnetization. $\mathbf{H}_{\text{anis}}(\mathbf{r}, t)$ takes into account the anisotropic nature of the material. This system of coupled partial differential equations almost always calls for numerical solutions. There are several software packages available for this task (Donahue and Porter 1999). The steady-state solution for larger (typically micron or bigger) bulk materials is a magnetic-domain structure.

3. Application to Nanomagnets: The Single-Domain Approximation

If the typical size of the nanomagnet is smaller than the single-domain limit, then the exchange field usually overwhelms the dipole field, thus forcing a parallel alignment of magnetic moments inside the particle. The single-domain limit for a cubic permalloy particle is about 50 nm. If this approximation is applicable, then the magnetization of a particle can be represented by a single vector (three scalar numbers), instead of a vector field. This results in a significant simplification for the qualitative understanding of the behavior of the particle, and the numerical simulation of dots becomes much easier. Moreover, the behavior of the single-domain particle becomes much more predictable, unlike the complex switching characteristics of a multi-domain particle with its complicated dependence on the ‘history’ of its magnetization. Single-domain particles have a very simple hysteresis loop, with a well defined switching field.

A nanomagnet logic device consists of a finite number of dots. Denoting the average magnetization and effective field of the i -th dot by $\mathbf{M}^{(i)}(t)$ and $\mathbf{H}_{\text{eff}}^{(i)}(t)$ respectively, the magnetization vector obeys the following equation of motion:

$$\frac{d\mathbf{M}^{(i)}(t)}{dt} = -\gamma \mathbf{M}^{(i)}(t) \times \mathbf{H}_{\text{eff}}^{(i)}(t) - \frac{\alpha\gamma}{M_s} [\mathbf{M}^{(i)}(t) \times \mathbf{M}^{(i)}(t) \times \mathbf{H}_{\text{eff}}^{(i)}(t)] \quad (3)$$

These are three ordinary differential equations for each dot. The effective field is given by:

$$\mathbf{H}_{\text{eff}}^{(i)}(t) = \mathbf{H}_{\text{ext}}^{(i)}(t) + \mathbf{N}^{(i)} \mathbf{M}^{(i)}(t) + \sum_{j \in \text{neighbors}} C_{ij} \mathbf{M}^{(j)}(t) \quad (4)$$

It consists of three parts: the first is the external field, the second is the self-demagnetization field which depends on the shape of the nanomagnet, and the third part describes the interaction with neighboring dots. $\mathbf{N}^{(i)}$ is a matrix containing three scalar numbers:

$$\mathbf{N}^{(i)} = \begin{bmatrix} N_x^{(i)} & 0 & 0 \\ 0 & N_y^{(i)} & 0 \\ 0 & 0 & N_z^{(i)} \end{bmatrix} \quad (5)$$

The above system of ODE-s is just one possible form of the single-domain approximation (or the so-called Stoner-Wohlfarth model (Hubert and Schafer 1998)). The matrices \mathbf{N} and \mathbf{C} are given for various approximations, see Hubert and Schafer (1998), Cowburn *et al.* (1999) and Stamps and Hillebrands (1999) for details.

There is still some debate in the literature under what conditions the single-domain approximation is valid, and when it breaks down. We performed most of our simulations both with and without the single-domain approximation, and ‘*a posteriori*’ verified its validity. Our calculations gave the same results (typically, with less than 10% deviation), however, the single-domain approximation often overestimates the switching fields. All the simulations used the material parameters for permalloy.

4. The Nanomagnet as Bistable Switch

Let us consider now a pillar-shaped single domain nanomagnet, as schematically shown in Fig. 1. Due to the strong shape anisotropy ($N_z < N_x, N_y$), the magnetization of a dot in steady state is always parallel with its longest axis, here pointing upwards or downwards.

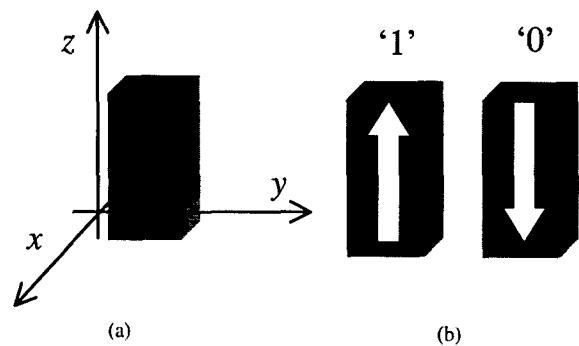


Figure 1. A single nanomagnet pillar (a) and its two stable magnetization states (b).

In logical sense, this dot is a bistable switch, and stores one bit of digital information. We assign the logical value '1' to the magnetization pointing up, and '0' for magnetization pointing down. The miniaturization of the magnets is ultimately restricted by the fact that their switching energies have to be larger than kT , and their switching speed is limited by the precession frequency. These restrictions still yield an impressive integration density of 10^{10} cm^{-2} , and speed in the GHz regime. Storage architectures address all dots individually by a read-write head or metallic wires.

5. Adiabatic Control of Magnetic Nanostructures: The Nanowire

A line of magnetically-coupled permalloy pillars can be thought of a magnetic nanowire.

Let us assume, that the leftmost pillar is pinned, i.e. its magnetization is fixed, and pointing up. We can consider the dipole chain as an inverter-chain in the logical sense (Fig. 2). If we can guarantee that the wire is in its ground state, it transmits binary information from its input to its output, and the state of every dot in the wire is determined by the input. Note that in the physical structure, the input dot is slightly thinner than the other dots, resulting in a higher switching field for this dot.

In order to move the wire from an arbitrary initial state to its ground state, we use adiabatic control by external fields. The details of this process are shown in Fig. 3. In the first phase, an external field is applied which is able to switch every dot, except for the input dot (which has a higher switching field, as discussed above). By the end of the first phase, the 'memory' of the structure is erased: the magnetic moments of the

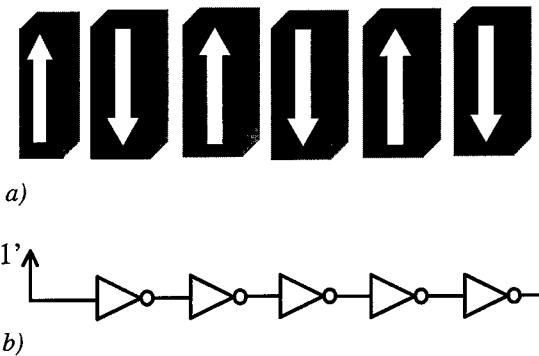


Figure 2. The physical structure of the inverter (a) and its logic equivalent (b).

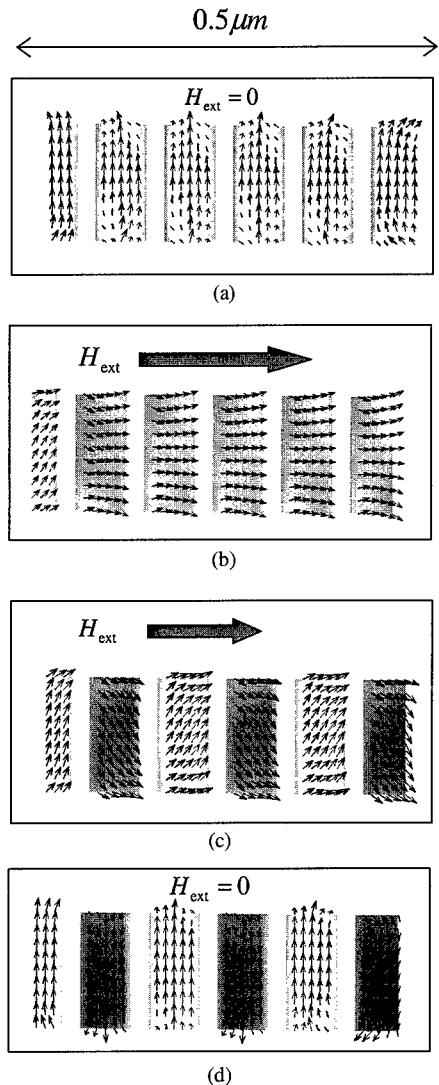


Figure 3. Adiabatic pumping of the nanowire. The initial metastable state (a) eliminated by a strong external field (b). Slowly releasing the field (c), the system relaxes to the zero-field ground state (d).

wire dots are in line with the strong external field, regardless of their previous state. In the second phase, this external field is slowly released. In this phase, the moments order according to the state of the first dot, which retained its magnetization. The term 'adiabatic control' is used since the dots always remain close to their ground state during the second phase. As no precession of the magnetization vector occurs, complicated nonlinear dynamics are suppressed. The results of the micromagnetic simulation (shown in Fig. 3) agree well with the qualitative expectations.

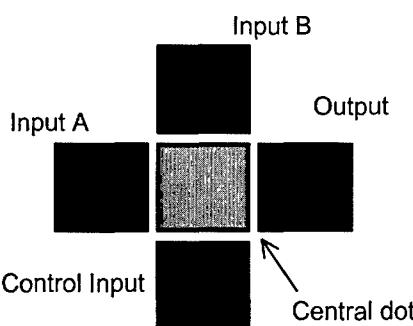


Figure 4. Schematic of the majority gate.

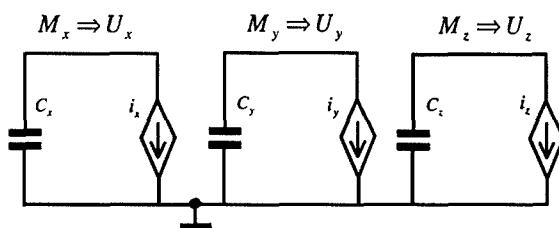


Figure 5. The equivalent-circuit model of a single nanomagnet dot.

The possibility of adiabatic control in magnetic nanostructures is one of the main results of our work. We think that even in large-scale arrays, not only the statistical properties of the system, but the magnetization state of individual dots can be reliably controlled.

6. Toward More Complex Structures: Majority Gate

The majority gate is a three-input, universal logic gate, which can realize basic logic functions, as schematically shown in Fig. 4. An OR gate is realized by setting its control input to '1' and it behaves as an 'AND' gate by setting the control input to '0'. The majority gate can be operated by the same 'adiabatic pumping' scheme as the wire. In the ground state, the central dot is antiparallel with the majority of input dots. Since the interaction between diagonally adjacent dots might have undesired effect on the operation of the device, the design requires more care than for a single wire.

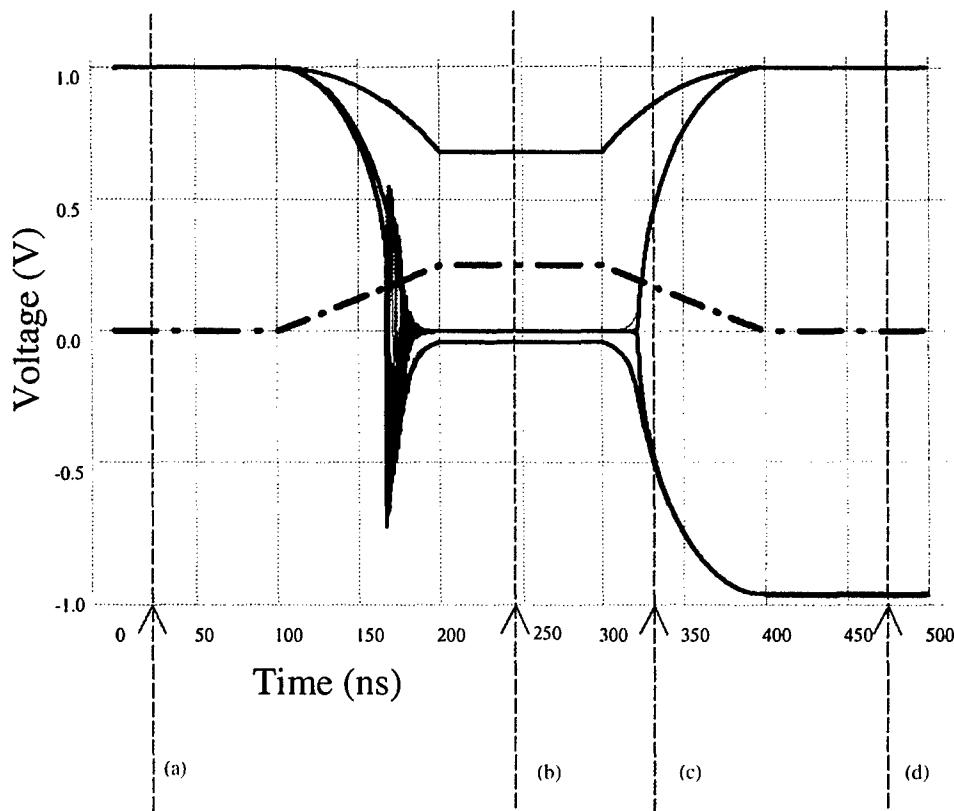


Figure 6. SPICE—simulation: the z -component of magnetization in a nanowire (in M_s units). The phases (a), (b), . . . corresponds to the stages in Fig. 3. The dashed line is the external control field.

The logic blocks presented above (inverting wire, majority gate), in principle, are sufficient for building a logic device of arbitrary complexity. Another degree of freedom is that nanomagnets with different shapes have a different response to external magnetic fields. Therefore, a homogenous external field can ‘target’ specific groups of similarly-shaped nanomagnets in a larger array. Our calculations show that we can take advantage of this fact for clocking the circuit by external fields, and realizing sequential circuits this way.

The realization of basic nanomagnet arrays appears feasible, given the technological base for data storage applications. The maximum possible complexity of nanomagnet networks mostly depends on the error tolerance of the design. Work on this issue is in progress.

7. Design Tools: The Circuit Model of Nanomagnets

An important result of our work is that nanomagnet logic devices can be designed with methods similar to electronic circuits. To fully exploit this analogy, we present a circuit model for nanomagnet dots.

We can identify the magnetization-vector components of each nanomagnet as a formal voltage, which obeys the differential Eqs. (3) and (4). Then, these equations are formally equivalent to the voltage change on a capacitor which is charged by current sources given by the right-hand sides of Eqs. (3) and (4). In this fashion, Eqs. (3) and (4) can formally be represented by a circuit, as schematically shown in Fig. 5.

Now we can use standard circuit-simulation tools, such as SPICE for micromagnetic design. Figure 6 shows the results from a particular nanowire-simulation. We can look at M_z as an electric signal in a real circuit.

8. Summary and Outlook

We proposed coupled ferromagnetic dots as possible building blocks for magnetic QCA-like nanocomputers. Carefully applied external fields make possible the individual, precise control of the state of nanomagnets inside a larger array. Starting from the micromagnetic equations, we have shown operating logic devices, and developed design tools for easily simulating them. Our work suggests that nanomagnets are promising candidates not only for nonvolatile information storage, but also for nanocomputing as well.

Acknowledgments

The authors would like to acknowledge helpful discussions with Prof. Árpád Csurgay Alexandra Imre. The work was supported in part by grants from the W.M. Keck Foundation and the Office of Naval Research.

References

- Cowburn R.P., Koltsov D.K., Adeyeye A.O., Welland M.E., and Tricker D.M. 1999. Physical Review Letters 83: 1042.
- Cowburn R.P. and Welland M.E. 2000. Science 287: 1466.
- Donahue M.J. and Porter D.G. 1999. OOMMF User’s Guide, Version 1.0, Interagency Report NISTIR 6376. National Institute of Standards and Technology, Gaithersburg, MD. A wealth of useful information about micromagnetic modeling is found at: <http://math.nist.gov/oommf/>.
- Hubert A. and Schafer R. 1998. Magnetic Domains. The Analysis of Magnetic Microstructures. Springer-Verlag, Berlin.
- Lent C.S., Tougaw P.D., Porod W., and Bernstein G.H. 1993. Nanotechnology 4: 49–57.
- Snider G.L., Orlov A.O., Amlani I., Zuo X., Bernstein G.H., Lent C.S., Merz J.L., and Porod W. 1999. J. Appl. Phys. 85: 4283.
- Stamps R.L. and Hillebrands B. 1999. Applied Physics Letters 75: 23.



Numerical Acceleration of Three-Dimensional Quantum Transport Method Using a Seven-Diagonal Pre-Conditioner

DAVID Z.-Y. TING

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

MING GU

Department of Mathematics, University of California, Los Angeles, CA 90024, USA; Department of Mathematics, University of California, Berkeley, CA 94720, USA

XUEBIN CHI

Department of Mathematics, University of California, Los Angeles, CA 90024, USA; R&D Center for Parallel Software, Institute of Software, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

JIANWEN CAO

R&D Center for Parallel Software, Institute of Software, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

Abstract. The open-boundary planar supercell stack method treats three-dimensional quantum transport in mesoscopic structures in a numerically stable and efficient manner. The method formulates quantum mechanical scattering problems for the supercell geometry as sparse linear systems, which can be solved by iterative methods. Recent improvement in the solution algorithm using a seven-diagonal pre-conditioner has resulted in over two orders of magnitude of numerical acceleration, bringing more flexibility in the range of problems that can be solved. We demonstrate the effectiveness of the seven-diagonal pre-conditioning algorithm by applying it to the studies of interface roughness in double barrier resonant tunneling structures and tunneling characteristics of n^+ poly-Si/SiO₂/p-Si with ultra-thin oxides undergoing dielectric breakdown.

Keywords: three-dimensional, quantum transport, sparse linear system, pre-conditioner

1. Introduction

The exactly solvable, real-space, open-boundary planar supercell stack method (OPSSM) has been applied to a variety of topics involving three-dimensional quantum transport in mesoscopic devices (Ting 1999a). The method was designed for studying elastic scattering effects due to impurities, interface roughness, and alloy disorder in our studies of 2D (double barrier heterostructures), 1D (quantum wires electron wave guides), and 0D (quantum dots) mesoscopic device structures. OPSSM studies have demonstrated:

interfacial inhomogeneities in double barrier resonant tunneling diodes can induce lateral localization of wave functions (Ting, Kirby and McGill 1994), strongly attractive impurities can produce additional transmission resonance (Kirby, Ting and McGill 1993), clustering effects in alloy barriers can reduce barrier effectiveness, and surface roughness in quantum dots can cause large fluctuations in transmission characteristics (Kirby, Ting and McGill 1994). In addition, OPSSM has also been used to study resonant tunneling via self-organized quantum dot states (Wang *et al.* 1998), and interface roughness effects (Ting 1998)

and dielectric breakdown (Ting 1999b) in n^+ poly-Si/SiO₂/p-Si tunnel structures containing ultra-thin oxide layers. In this paper, we present the results of our recent work on numerical acceleration. These new implementations have led to up to two orders of magnitude gains in computational efficiency, greatly extending the range of applicability of OPSSM.

2. Method

The generic device structure treated by OPSSM consists of a slab of active region sandwiched between two semi-infinite, homogeneous, flat-band electrode regions. The active region is described by a stack of planar supercells that can take on quasi-3D variations. Let the z -axis be the direction of current flow. Then the active region is composed of a stack of N_z layers perpendicular to the z -direction, with each layer containing a periodic array of rectangular planar supercells of N_x by N_y sites. Within each planar supercell, the potential assumes lateral variations as dictated by device geometry. A one-band nearest-neighbor tight-binding Hamiltonian is used to describe the potential and effective mass variations over this volume of interest (Ting 1999a).

OPSSM formulates quantum mechanical scattering problems for supercell geometries as a complex, non-Hermitian, sparse linear system $Ax = b$. Here A is the Hamiltonian matrix, augmented by special terms representing opening boundary conditions; x is a vector containing the coefficients of the tight-binding orbitals (which is equivalent to the envelope function in the effective mass approximation); and b is a vector describing the open boundary condition (Ting 1999a). As the order of the linear system ($N_x \times N_y \times N_z$) can be quite large, direct solution methods are impractical since the LU factorization of A is almost dense, requiring computational times that are proportional to $N_x^3 \times N_y^3 \times N_z$, which is prohibitively expensive for all but the smallest supercells. However, the linear system can be solved iteratively using the Quasi-Minimal Residual (QMR) method (Freund and Nachtigal 1991, 1996), which is capable of treating non-Hermitian systems. As with other iterative methods, the convergence behavior and efficiency of QMR depends on the eigenvalue spectrum of A . In some cases QMR may not work or may converge too slowly to be practical. These problems can be greatly ameliorated by employing a good pre-conditioner (Freund and Nachtigal

1991, 1996). The idea is to use a pre-conditioning matrix M to transform the original problem into an equivalent one $(M^{-1}A)x = M^{-1}b$. The transformed linear system has the same solution as before, but can have a completely different eigenvalue spectrum, which could be much easier to solve than the original problem. The ideal pre-conditioner M should be easily invertable, and should approximate A in some sense. (Note that if $M = A$, the problem is solved.) While the inversion of M introduces additional computation, for good choices of M , however, the transformed problem can be solved in significantly less number of iterations than the original one. While there are general techniques, such as the incomplete LU factorization (Meijerink and van der Vorst 1997, Chan and van der Vorst 1997), for generating pre-conditioners for general sparse matrices, there is still a practical need for pre-conditioning techniques that can take advantage of the specific features of a particular problem class. Obviously, the more knowledge we have of the special features, the better we could exploit it to in the construction of a good pre-conditioner. In this work, we present the results from the application of a seven-diagonal pre-conditioner (Cao unpublished), which appears to be well suited to our system.

The design of seven-diagonal pre-conditioner is based the structure of the sparse matrix A , which we briefly describe here and refer reader to Ting (1999a) for details. There are several ways the non-zeros are introduced in the matrix A . Intra planar supercell interactions (on-site and nearest neighbor) give rise to five non-zero diagonals in A . Interactions connecting adjacent planar supercells contribute to two diagonals. In-plane supercell periodicity generates four sparsely populated diagonals. And finally, the open boundary conditions introduce two dense blocks near the upper left and lower right corners of A . To achieve both effectiveness and efficiency, we choose to ignore the dense blocks in constructing a pre-conditioner. Taking account of these dense blocks in M could make M harder to factorize. Because they arise from open boundary conditions associated with incident, reflected and transmitted plane waves, rather than the mesoscopic structures, we assume that they do not affect the eigenvalues of A as much as the rest of the non-zeros. Due to the nature of the open boundary condition, the dense blocks are related to discrete Fourier transform. They are well-conditioned, and do not need to be explicitly generated in order to perform the matrix-vector multiplication $A v$ for any given vector v , which is an

operation required at each iteration step of the QMR solver. The inclusion of the dense blocks in \mathbf{M} would also have made it more costly to multiply \mathbf{M}^{-1} to arbitrary vectors. The pre-conditioner is thus constructed from \mathbf{A} by ignoring the dense blocks and the four sparse diagonals generated by planar supercell periodicity. Effectively, \mathbf{M} represents a 7-diagonal pre-conditioner of \mathbf{A} , even though \mathbf{A} itself has many more non-zeros in addition to these 7 diagonals. We set $\mathbf{M} = \mathbf{L}\mathbf{U}$, where \mathbf{L} and \mathbf{U} are lower and upper triangular matrices, respectively, so that $\mathbf{L} + \mathbf{U}$ has the same non-zero pattern as \mathbf{A} except those dense blocks. The non-zeros of \mathbf{L} and \mathbf{U} are generated in the way that mimics the in-complete LU factorization of \mathbf{A} sans dense blocks and sparse diagonals. The application of this pre-conditioner turns out to be very effective, as shown in the next section.

3. Results and Discussions

Figure 1 illustrate the performance gains achieved using a 7-diagonal pre-conditioner in typical applications involving the studies of interface roughness effects in GaAs/AlAs double barrier resonant tunneling heterostructures (Ting, Kirby and McGill 1994). The supercell stack consists of $N_z = 26$ layers, and the size of the planar supercell is 64×64 . We plot the computed transmission coefficient and the number of iterations required for reaching convergence over an energy range of interest, using both non-pre-conditioned and pre-conditioned algorithms, and with different convergence tolerances. Cutoffs of 10,000 and 50,000 iterations are imposed to limit the total run time. In general, we find that the number of iterations required to reach convergence increases with increasing incoming electron energy, or decreasing deBroglie wavelength. Without pre-conditioning, almost all points in the spectrum with energy greater than the resonance energy fail to converge within 50,000 iterations. The application of the 7-diagonal pre-conditioner substantially reduces the number of iterations required to reach convergence. As a result, very few points in the spectrum failed to converge. Figure 2 shows application to the study of ultra-thin oxide tunnel structures with embedded nano-filaments as a model of dielectric breakdown (Ting 1999b). The supercell stack consists of $N_z = 37$ layers of 64×64 planar supercells. In this case, convergence was reach within the limit of 60,000 iterations for all cases. However, without pre-conditioning, more than half of points

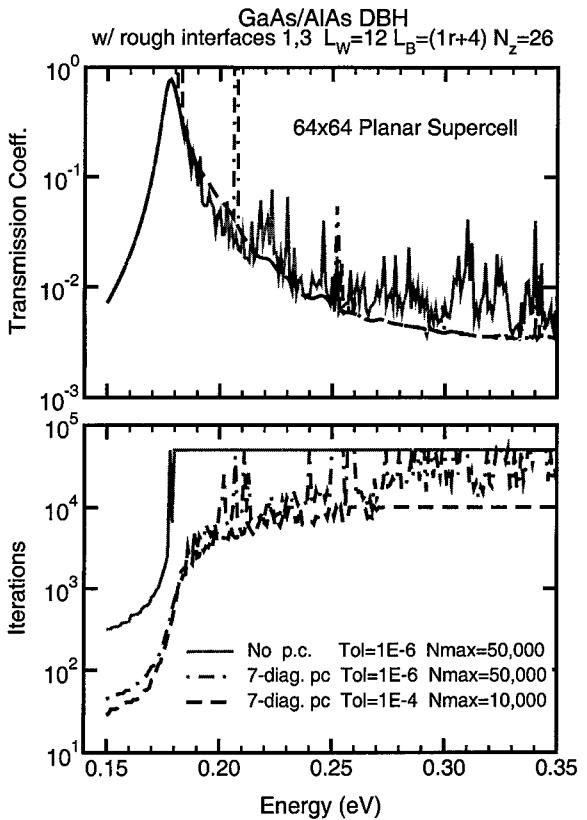


Figure 1. Tunneling calculation for a GaAs/AlAs double barrier structure with interface roughness are performed with and without preconditioning, using different convergence tolerances and limits on iteration numbers. The upper panel shows transmission coefficient spectra, and the lower panel shows convergence iteration counts.

required 60,000 iterations. With pre-conditioning, all except for 2 points converged in fewer than 400 iterations. Table 1 summarizes the timing results for some typical applications, and documents the degree of numerical acceleration resulting from the application of the 7-diagonal pre-conditioner. The combination of a suitable pre-conditioner and a more judicious choice of convergence tolerance resulted in almost a 300-fold speed-up in large supercell applications.

With over two orders of magnitude of numerical acceleration, we are now able to obtain more accurate results by using larger supercells. Previously, calculations of interface roughness effects on resonant tunneling properties of double barrier heterostructures were performed using modestly sized (20×20 or 32×32) planar supercells (Ting, Kirby and McGill 1994, Ting and McGill 1996). The resulting transmission spectra always contained satellite peaks above the main resonance. While these satellite peaks were very useful

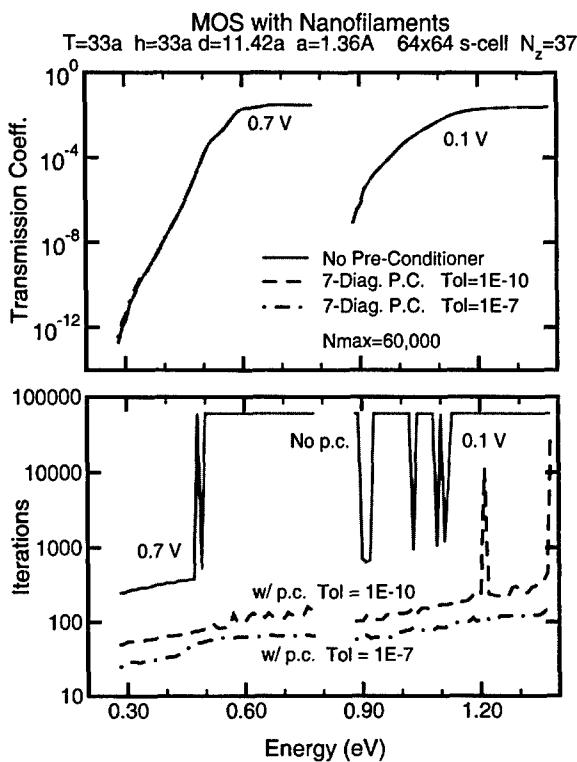


Figure 2. Tunneling calculation for an ultra-thin oxide layer embedded with nano-filaments under two biasing conditions are performed with and without preconditioning, using different convergence tolerances and limits on iteration numbers. The upper panel shows transmission coefficient spectra, and the lower panel shows convergence iteration counts.

in clarifying the interface roughness assisted resonant tunneling mechanism, they are also identified as artifacts induced by the periodic boundary condition associated with the finite supercell (Ting, Kirby and McGill 1994). Ting, Kirby and McGill (1994) suggested that the use of larger supercells could suppress these artifacts. However, previous attempts plagued by convergence problems, as is evident from the

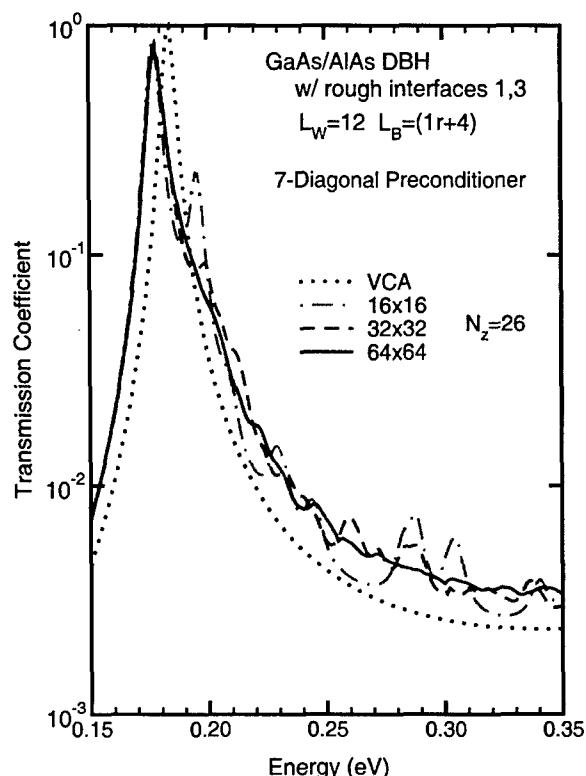


Figure 3. Transmission coefficient spectra of GaAs/AlAs double barrier structure computed using various supercell sizes. Numerical acceleration by pre-conditioning made large supercell calculations possible, thus enabling the reduction of artifacts due to supercell-periodic scattering.

non-preconditioned results in Fig. 1, which shows that the 64×64 results are not usable due to lack of convergence. Figure 3 shows transmission coefficients calculated for three different planar supercell sizes (16×16 , 32×32 , and 64×64) using the 7-diagonal pre-conditioner. The finite supercell artifacts in the 16×16 and 32×32 results have essentially disappeared from the 64×64 result.

Table 1. Speed-up obtained in typical 3D scattering calculations with a 7-diagonal pre-conditioner.

	No pre-conditioner	7-Diagonal preconditioner	Speed-up
DBH $16 \times 16 \times 26$	8.92 h	1.66 h	5.4
DBH $32 \times 32 \times 26$	208.7 h	17.6 h	12
DBH $64 \times 64 \times 26$	1511 h (Poor convergence)	283.2 h (Good convergence)	$\gg 5$
MOS $64 \times 64 \times 37$ (Tol = 1E-10)	1618.5 h	23.2 h	70
MOS $64 \times 64 \times 37$ (Tol = 1E-10)	1618.5 h (Tol = 1E-10)	5.7 h (Tol = 1E-7)	284

In summary, we demonstrated over two orders of magnitude of numerical acceleration in our solution algorithm of 3D quantum mechanical scattering code by using a seven-diagonal pre-conditioner. The improvement brings more flexibility in the range of quantum device modeling problems that can be solved numerically.

Acknowledgment

A part of the work described in this paper was performed by DZT at the Jet Propulsion Laboratory (JPL), California Institute of Technology, and was sponsored by the Defense Advanced Research Projects Agency through a JPL Technology Affiliates Program with the HRL Laboratories. MG acknowledges support from HRL Laboratories and the State of California under a MICRO grant (UC MICRO 99-050). The authors thank J. N. Schulman and R. Caflisch for helpful discussions.

References

- Cao J.W. unpublished.
- Chan T.F. and van der Vorst H. 1997. In: Keyes D.E., Samed A., and Venkatakrishnan V. (Eds.), Parallel Numerical Algorithms, ICASE/LaRC Interdisciplinary Series in Science and Engineering, Vol. 4, Kluwer Academic, Dordrecht, pp. 167–202.
- Freund R.W. and Nachtigal, N.M. 1991. *Numer. Math.* 60: 315.
- Freund R.W. and Nachtigal N.M. 1996. *ACM T Math Software* 22: 46.
- Kirby S.K., Ting D.Z.-Y., and McGill, T.C. 1993. *Phys. Rev. B* 48: 15237.
- Kirby S.K., Ting D.Z.-Y., and McGill T.C. 1994. *Semicond. Sci. Tech.* 9(Suppl): S918.
- Meijerink J.A. and van der Vorst H. 1997. *Math. Comp.* 31: 148.
- Ting D.Z.-Y. 1998. *Appl. Phys. Lett.* 73: 2769.
- Ting D.Z.-Y. 1999a. *Microelectronics J.* 30: 985.
- Ting D.Z.-Y. 1999b. *Appl. Phys. Lett.* 74: 585.
- Ting D.Z.-Y., Kirby S.K., and McGill T.C. 1994. *Appl. Phys. Lett.* 64: 2004.
- Ting D.Z.-Y. and McGill T.C. 1996. *J. Vac. Sci. Technol. B* 14: 2790.
- Wang J.N., Li R.G., Wang Y.Q., Ge W.K., and Ting D.Z.-Y. 1998. *Microelectron Eng.* 43-4: 341.



Numerical Investigation of Shot Noise between the Ballistic and the Diffusive Regime

M. MACUCCI, G. IANNACCONE AND B. PELLEGRINI

Dipartimento di Ingegneria dell'Informazione, Università degli studi di Pisa, Via Diotisalvi 2, I-56126 Pisa, Italy

Abstract. We investigate shot noise suppression in several mesoscopic structures by means of a numerical approach based on the computation of the transmission matrix with the recursive Green's function method. We retrieve the “universal” values of the suppression factor obtained with random matrix theory for chaotic cavities and diffusive conductors. We then extend the investigation to more complex structures, such as multiple cascaded cavities and partially diffusive systems, and discuss the consequences on the shot noise suppression factor. Finally, we analyze the behavior of shot noise in an electron waveguide containing a large number of scatterers as the spatial position of the scatterers is changed from a regular array to a random distribution.

Keywords: shot noise, mesoscopic, chaos, ballistic

1. Introduction

During the last few years remarkable theoretical (Lesovik 1989, Büttiker 1990, Beenakker and Büttiker 1992, Jalabert, Pichard and Beenakker 1994, González *et al.* 1998) and experimental (Kumar *et al.* 1996, Liefrink *et al.* 1994, Oberholzer *et al.* 2001) results have drawn significant attention to the issue of shot noise suppression in mesoscopic conductors. The most recent theoretical work in this field has been based on the random matrix approach (RMT), which has allowed prediction of the shot noise suppression down to 1/3 of the full shot value in diffusive conductors (Beenakker and Büttiker 1992) and of the suppression down to 1/4 in chaotic ballistic cavities (Jalabert, Pichard and Beenakker 1994). The RMT approach is quite powerful, but it cannot be easily extended to generic geometries; we have been interested in expanding the investigation of shot noise suppression to arbitrary mesoscopic structures, and, to this purpose, we have developed a numerical method based on an optimized recursive Green's function technique. With this method, we can treat generic structures, with the inclusion of the effects of atomistic distributions of dopants leading to a diffusive regime, and we can handle situations with a few hundreds of propagating modes. It is

possible to show that the “universal” suppression factors 1/3 and 1/4 are easily retrieved, respectively, for a conductor with a large enough density of elastic scatterers and for a structure with a symmetric cavity with small enough input and output apertures. We study shot noise in nanostructures containing single and multiple cascaded cavities, noticing that the shot noise suppression is substantially independent of the number of cavities, and then take into consideration the case in which one of the cavities is filled with randomly distributed scatterers, arguing, on the basis of a simple circuit analogy, why the shot noise reduction factor becomes the same as for purely diffusive conductors. Finally, we investigate the transition that shot noise suppression in an electron waveguide containing a large number of scatterers undergoes as we move from a regular spatial distribution of such scatterers to a random distribution.

2. Model

Although our approach is general and can be applied to an arbitrary potential landscape, we consider, for the sake of computational simplicity, a device geometry defined by hard walls, with obstacles and boundaries characterized by right angles. The transmission

matrix t , whose elements represent the transmission coefficient from each input mode to each output mode, is computed by means of the recursive Green's function approach (Sols *et al.* 1989, Macucci, Galick and Ravaioli 1995), which has been specifically optimized to guarantee sufficient numerical precision when handling up to a thousand of the slices characterized by constant transverse potential into which the structure has to be subdivided for the calculations that we will be presenting. Once t has been obtained, we compute the transmission coefficients in a representation in which the transmission matrix is diagonal, multiplying t by its hermitian conjugate t^\dagger and finding the eigenvalues T_i of tt^\dagger . Following Lesovik (1989) the shot noise power density can be written as

$$S_I = 4 \frac{q^2}{h} |qV| \sum_i T_i (1 - T_i), \quad (1)$$

where h is the Planck constant, q is the electron charge and V is the applied voltage. Since the power spectral density of full shot noise is

$$S_{I_f} = 2qI = 2q \frac{2q^2}{h} |V| \sum_i T_i, \quad (2)$$

we can conclude that the Fano factor γ , i.e. the ratio of the actual shot noise power spectral density to the full shot noise, is given by

$$\gamma = \frac{\sum_i T_i (1 - T_i)}{\sum_i T_i}, \quad (3)$$

which can be immediately evaluated once the T_i coefficients are known.

3. Numerical Results

We have first investigated the shot noise suppression in chaotic cavities (defined by apertures that are much narrower than the cavities themselves), retrieving (Macucci, Iannaccone and Pellegrini 2001) the value of 1/4 for the Fano factor, as predicted by Jalabert *et al.* (1994), if the number of propagating modes is larger than about 20. We have then studied a more complex structure, made up of two cascaded cavities, each 5 μm long, created in an electron waveguide with a width of 5 μm by delimiting them with diaphragms 250 nm thick and 1 μm wide, as shown in the inset of Fig. 1. We report the Fano factor for this structure in Fig. 1 as

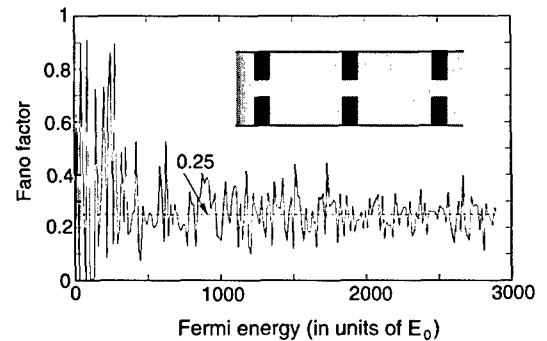


Figure 1. Fano factor for two cascaded chaotic cavities as a function of the Fermi energy, expressed in units of the threshold E_0 for propagation of the lowest mode in the empty waveguide. The inset contains a graphic representation of the confinement potential.

a function of the Fermi energy (expressed as a multiple of the threshold energy E_0 for propagation of the lowest mode in the empty waveguide), and notice that the average value is around 0.25, as in the case of a single cavity. The structures we are studying are relatively large, in order to allow propagation of a sufficiently large number of modes, to be in the regime in which the “universal” suppression factors are meaningful (Beenakker and Büttiker 1992).

We have also computed the Fano factor for three cascaded cavities, obtaining results that, although with larger fluctuations, are almost coincident with those for two cavities. The same happens for larger numbers of cascaded cavities, and even if we include intermediate diaphragms with different widths, as long as the rightmost and the leftmost apertures are symmetric. We notice that the actual shot noise suppression factor fluctuates rather widely as a function of the Fermi energy for all of the numerical results, and equals the asymptotic value predicted by random matrix theory only on the average.

A qualitatively different behavior is however observed if at least one of the cascaded cavities is filled with randomly distributed obstacles, which lead to a complex scattering pattern and to transport in the diffusive regime, i.e. a condition in which the elastic mean free path is much smaller than the device dimensions. In Fig. 2 we report the noise power spectral density as a function of the Fermi energy for two cascaded cavities, each with a length of 5 μm and a width of 5 μm , delimited by constrictions that are 1 μm wide and 0.25 μm long. Within the cavity region we have included 200 randomly distributed hard-wall 56.2 nm \times 50 nm obstacles. Although the cavity is delimited by symmetric apertures, the Fano factor moves up to slightly less

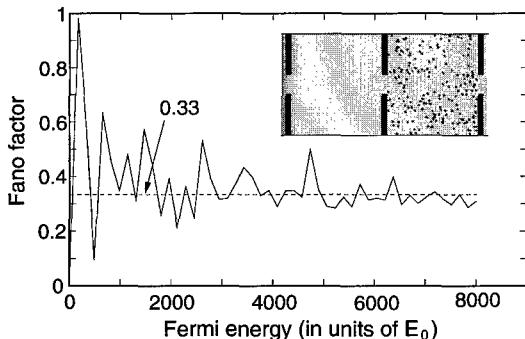


Figure 2. Fano factor for two cascaded chaotic cavities, one of which is filled with randomly distributed scatterers, as a function of the Fermi energy, expressed in units of the threshold E_0 for propagation of the lowest mode in the empty waveguide. The inset contains a graphic representation of the confinement potential.

than $1/3$, significantly departing from the $1/4$ result and reaching a typically diffusive behavior. The inset in the figure shows the device geometry, with the position of the obstacles.

An extremely simplified interpretation of this behavior can be derived from a circuit analogy. Let us consider a series of two current noise sources, with power spectral densities S_{I_1} and S_{I_2} , providing contributions of the same order of magnitude (as they are both of shot origin and share the same average current) and associated with different resistances R_1 and R_2 , each of which is in parallel with the corresponding current noise source. If we want to determine the current noise power spectral density $S_{I_{out}}$ they produce on an external load R , we obtain $S_{I_{out}} = (S_{I_1}R_1^2 + S_{I_2}R_2^2)/(R_1 + R_2 + R)^2$, therefore the predominant contribution is the one associated with the larger resistance, which in our case corresponds to the diffusive region. Clearly, this is not an exact analogy, because the electron waveguide sections do not rigorously correspond to circuit elements in series, although the presence of a diffusive region has a strongly decoupling action between the different sections.

Another interesting aspect of the transition from ballistic to diffusive transport can be observed by applying our computational method to a quantum wire containing scatterers and looking at the dependence of the shot noise suppression factor on the position of such scatterers. If we have a regular pattern of scatterers, arranged in a square lattice, it has been shown (Macucci in press) that, at least for relatively small numbers of scatterers, shot noise is suppressed by a factor increasing with the portion of the waveguide surface occupied by the scatterers and saturating around 0.16. On the other hand, we

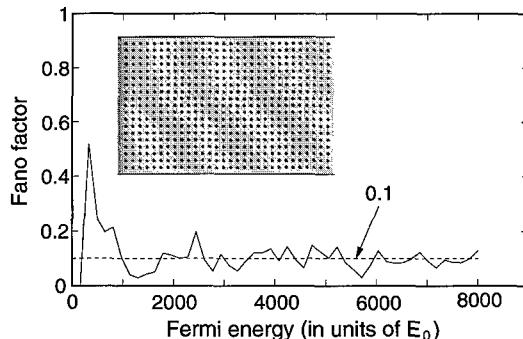


Figure 3. Fano factor for an electron waveguide filled with a square lattice of scatterers, as a function of the Fermi energy, expressed in units of the threshold E_0 for propagation of the lowest mode in the empty waveguide. The inset contains a graphic representation of the confinement potential.

know that, for a large number of randomly positioned scatterers, shot noise is suppressed by the universal factor $1/3$ (Macucci, Iannaccone and Pellegrini 1999). We have performed a calculation of the Fano factor for a section of electron waveguide containing a very large number of square obstacles (570), each with a side 200 times smaller than the waveguide width, for two cases differing for the spatial arrangement of the scatterers, but not for their density. In one case we have a regular square lattice, with 19 rows and 30 columns, in the other case we generate the coordinates of the scatterers as randomly distributed variables over the same region of space. Results are shown in Fig. 3 (for the square lattice) and in Fig. 4 (for the random case), in which we report the Fano factor as a function of the Fermi energy, expressed as a multiple of the energy for propagation of the lowest mode in the empty waveguide. Each figure

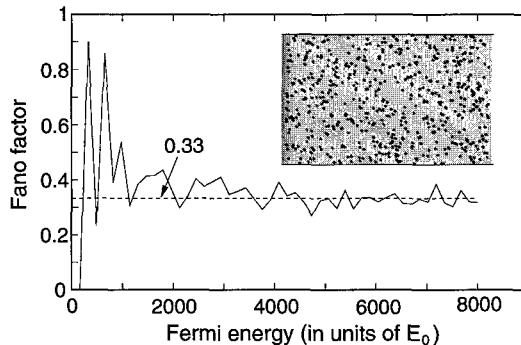


Figure 4. Fano factor for an electron waveguide filled with randomly distributed scatterers, as a function of the Fermi energy, expressed in units of the threshold E_0 for propagation of the lowest mode in the empty waveguide. The inset contains a graphic representation of the confinement potential.

contains an inset showing the position of the obstacles within the waveguide. It is apparent that, although the density of scatterers is the same in the two cases, the noise suppression sharply differs: for the regular lattice we observe an average value of the Fano factor around 0.1, which, considering the relatively low scatterer-to-waveguide area ratio, is in good agreement with the results obtained in Macucci (in press); when, instead, scatterers are distributed randomly, the 1/3 “universal” suppression factor predicted by random matrix theory (Beenakker and Büttiker 1992) is immediately retrieved.

4. Conclusions

We have investigated shot noise suppression in mesoscopic conductors in a regime that varies from ballistic, with the inclusion of simple scattering geometries, to diffusive, observing how the shot noise suppression factor varies and fluctuates around the “universal” values 1/4 and 1/3 for the chaotic cavities and for the diffusive regime, respectively. We have also observed that the 1/4 suppression factor is not influenced significantly by the characteristics and number of cascaded chaotic cavities, as long as the leftmost and rightmost apertures are of the same width. Furthermore, we have shown that the presence of a diffusive region within an electron waveguide leads to a Fano factor around 1/3 with little influence from the other geometrical details of the structure, and we have justified this result on the basis of a simple circuit analogy. Finally, we have investigated the change in the shot noise suppression factor in an electron waveguide, as the position of a large number of scatterers is varied from regular to random without varying their spatial density; a transition is observed from transport in a periodic structure to

the diffusive regime. Further work is planned to better understand this transition as the scatterer arrangement is gradually changed from regular to random and as a function of the actual statistical distributions used for the scatterer coordinates.

Acknowledgments

We acknowledge financial support from the Italian National Research Council (project 5% Nanotecnologie).

References

- Beenakker C.W.J. and Büttiker M. 1992. Phys. Rev. B 46: 1889.
- Büttiker M. 1990. Phys. Rev. Lett. 65: 2901.
- González T., González C., Mateos J., Pardo D., Reggiani L., Bulashenko M., and Rubí J.E. 1998. Phys. Rev. Lett. 80: 2901.
- Jalabert R.A., Pichard J.-L., and Beenakker C.W.J. 1994. Europhys. Lett. 27: 255.
- Kumar A., Saminadayar L., Glattli D.C., Jin Y., and Etienne B. 1996. Phys. Rev. Lett. 76: 2778.
- Lesovik G.B. 1989. Pis'Ma V Zhurnal Eksperimental'Noi i Teoreticheskoi Fiziki 49: 513.
- Liefrink F., Dijkhuis J.I., de Jong M.J.M., Molenkamp L.W., and van Houten H. 1994. Phys. Rev. B 49: 14066.
- Macucci M. 2002. Shot noise suppression due to an antidot lattice. Physica B 314: 494.
- Macucci M., Galick A., and Ravaioli U. 1995. Phys. Rev. B 52: 5210.
- Macucci M., Iannaccone G., and Pellegrini B. 1999. Quantum-mechanical simulation of shot noise in the elastic diffusive regime. In: Proceedings of the 15th International Conference on Noise in Physical Systems and 1/f Fluctuations. Hong Kong. p. 325.
- Macucci M., Iannaccone G., and Pellegrini B. 2001. Shot noise suppression in single and multiple ballistic and diffusive cavities. In: Proceedings of ICNF 2001, Gainesville, FL.
- Oberholzer S., Sukhorukov E.V., Strunk C., Schönenberger C., Heinzl T., and Holland M. 2001. Phys. Rev. Lett. 86: 2114.
- Sols F., Macucci M., Ravaioli U., and Hess K. 1989. J. Appl. Phys. 66(8): 3892.



On Ohmic Boundary Conditions for Density-Gradient Theory

M.G. ANCONA

Naval Research Laboratory, Washington, DC, USA

ancona@estd.nrl.navy.mil

D. YERGEAU AND Z. YU

Stanford University, Palo Alto, CA, USA

B.A. BIEGEL

NASA Ames Research Center, Moffett Field, CA, USA

Abstract. Conventional ohmic boundary conditions are shown to be inconsistent with density-gradient (DG) theory. New ohmic conditions that are consistent with DG theory are then derived and illustrated with two device examples. The first example uses a short p-n diode to understand the basic situation while the second treats a MOSFET contact and studies the “insulator proximity effect” seen at the point/edge where the ohmic contact abuts an insulator.

Keywords: density-gradient, ohmic contacts, boundary conditions

1. Introduction

Density-gradient (DG) theory is a well-known generalization of diffusion-drift (DD) theory that enables lowest-order effects of quantum mechanics to be included in conventional device simulations (Ancona and Tiersten 1987, Ancona and Iafrate 1989, Ancona 1990a). This theory has been applied to a variety of device problems including inversion layer (Ancona 2000), SOI (Wettstein, Schenk and Fichtner 2001) and heterostructure (Ancona 1990b) confinement, random impurity effects in MOSFETs (Asenov *et al.* 2001) and tunneling from semiconductors (Ancona *et al.* 2000) and metals (Ancona 1992).

In steady-state the equations of DG theory governing electron and hole transport inside a semiconductor, as expressed in terms of quasi-Fermi level variables, are

$$\nabla \cdot (\mu_n n \nabla \Phi_n) = -R, \quad \nabla \cdot (\mu_p p \nabla \Phi_p) = R \quad (1a)$$

$$\nabla \cdot [b_n \nabla s] = \frac{s}{2}(\Phi_n + \phi_{ho} - \psi),$$

$$\nabla \cdot [b_p \nabla r] = \frac{r}{2}(-\Phi_p + \phi_{po} + \psi) \quad (1b)$$

$$\nabla \cdot (\varepsilon_s \nabla \psi) = q(n - p + N_a - N_d) \quad (1c)$$

where $s^2 = n$, $r^2 = p$, Φ_n and Φ_p are the respective quasi-Fermi levels for electrons and holes, $\phi_{ho}(n)$ and $\phi_{po}(p)$ are the density-dependent parts of the electron and hole chemical potentials (which typically take either Maxwell-Boltzmann or Fermi-Dirac form), b_n and b_p measure the strengths of the gradient (quantum) contributions to the electron and hole chemical potentials and all other quantities have their usual meanings. These governing equations of DG theory are 5 PDEs for the 5 variables ψ , n , p , Φ_n and Φ_p . With appropriate boundary conditions appended, they can be solved in order to analyze a variety of device situations involving quantum effects.

Of particular interest for this paper are the boundary conditions used to represent ohmic contacts. Ohmic BCs are peculiar in that their physical fidelity is almost always unimportant. Instead of attempting to simulate

the complex physics (including tunneling) of a low-resistance metal-semiconductor junction, one is instead generally satisfied merely to have simple, easily implemented conditions that give good numerical behavior with little added contact resistance. In DD theory—which obtains from (1) when b_n and b_p vanish—the usual conditions on the electron and hole quasi-Fermi levels are

$$\Phi_n = \Phi_p = V_0 \quad (2a)$$

where V_0 is the applied voltage. Their equality implies interfacial equilibrium with

$$\begin{aligned} \psi &= V_0 - \phi_{po}(p) = V_0 + \phi_{no}(n), \\ n &= n_{eq}, \quad p = p_{eq}, \end{aligned} \quad (2b)$$

of which conditions only one (on ψ) is needed to solve DD boundary value problems.

2. Ohmic Boundary Conditions for DG Theory

The naïve approach for handling an ohmic contact in DG theory is simply to impose the DD conditions (2). That this fails to work under all conditions is the main point of this paper.

To derive consistent BCs, the standard field theoretic approach is to employ integral forms of the governing equations. To illustrate, the integral form of the force balance equation on electrons is

$$\int_S \mathbf{n} \cdot \Phi_n dS = \int_V \mathbf{f}_n dV \quad (3a)$$

where V is an arbitrary volume with surface S , \mathbf{n} is the outward normal vector on S and \mathbf{f}_n is the (drag) force exerted by the lattice on the electrons (which in the bulk is given by \mathbf{v}_n/μ_n). The BC associated with (3a) is then derived by choosing V to be a Gaussian pillbox bisected by the semiconductor-metal interface and allowing the pillbox to collapse to the interface. Doing so, we find that Φ_n must be continuous if the interface exerts no force on the electrons (as should be the case for an ohmic contact) so that the right hand side of (3a) gives no contribution in the limit. Employing a similar argument for holes, we find that Φ_n is also continuous across the ohmic interface and thus have shown that the DD ohmic BCs of (2a) are consistent with DG theory.

Proceeding in a precisely analogous fashion, we next derive a consistent condition on the electron density for DG theory by starting from the integral form of (1b):

$$\int_S \mathbf{n} \cdot [b_n \nabla s] dS = \int_V \frac{s}{2} (\Phi_n - \phi_{no} - \psi) dV \quad (3b)$$

Again applying this to a Gaussian pillbox and taking the appropriate limits, we obtain

$$\mathbf{n} \cdot \nabla s = 0 \quad (4a)$$

where we have assumed the metal is ideal so that its carrier density is uniform and that the electron density in the semiconductor remains bounded as the limit is taken. A similar Neumann condition may also be derived for the holes

$$\mathbf{n} \cdot \nabla r = 0 \quad (4b)$$

These derivations show that for consistency with DG theory the BCs on the carrier densities at an ohmic contact should be Neumann rather than Dirichlet as in the DD BCs in (2b).

Now it is true that under most circumstances (see below) the electrostatics will constrain the majority carrier density—say, holes—forcing quasi-charge neutrality with $p \cong N_A$. This implies that if one uses the DD ohmic BC on p in (2b), (4b) will still be satisfied and little error will accrue. However, the electrostatics imposes no such constraints on the minority carriers (electrons) and so enforcing the DD condition $n = n_{eq}$ will lead to a violation of (4a) whenever DG (quantum) effects are significant near the contact. This inconsistency is also evident in the fact that in DG theory $n_{eq} \neq n_i^2/p_{eq}$.

Finally, in DD theory the conditions on ψ in (2b) come directly from (2a) and the definitions of Φ_n and Φ_p .¹ In DG theory neither of these conditions will be strictly valid because they are absent the gradient corrections demanded by (1b). Nevertheless, the condition on ψ derived from the majority carriers (holes), i.e., $\psi = V - \phi_{po}(p)$, will usually be well satisfied because again the electrostatics forces the majority carrier density to vary slowly.

3. Example: Short pn Diode

If the minority carrier diffusion length in the region contacted ohmically is comparable to the size of the region, then minority carrier current will be significant

and the proper handling of the ohmic BC for minority carriers will be critical. We therefore examine the case of a short pn diode, i.e., one for which $L < \sqrt{D\tau}$, using a general-purpose 1-D DG solver. In Fig. 1 we compare the density profiles computed using DG theory with conventional ohmic BCs and with the new DG BCs. The majority carrier profiles in Fig. 1(a) are grid independent and, as expected, are nearly identical apart from the slope at the contact. For the minority carriers

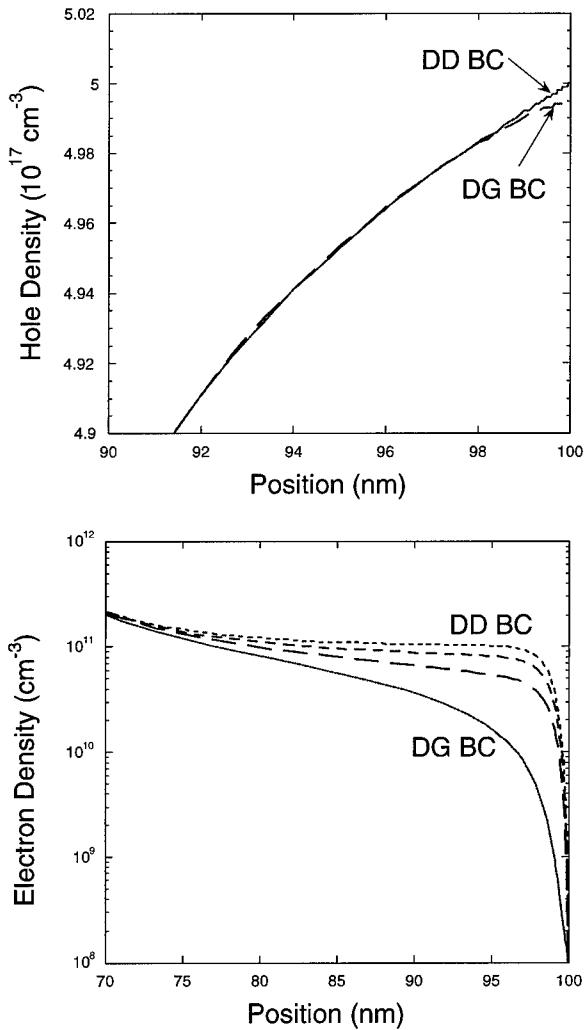


Figure 1. (a) Profiles of the hole density in the p-type region of a short pn diode showing good agreement in the treatment of the majority carrier. The only significant difference is due to the DG BC enforcing zero slope at the contact located at $x = 100$ nm. (b) Profiles of the electron density in the p-type region of a short pn diode (contact located at $x = 100$ nm) showing the inconsistency of the DD BCs in their treatment of the minority carriers. The several DD BC (dashed) curves differ only in the grids used in the calculations. The DG BC instead gives the grid-independent result shown.

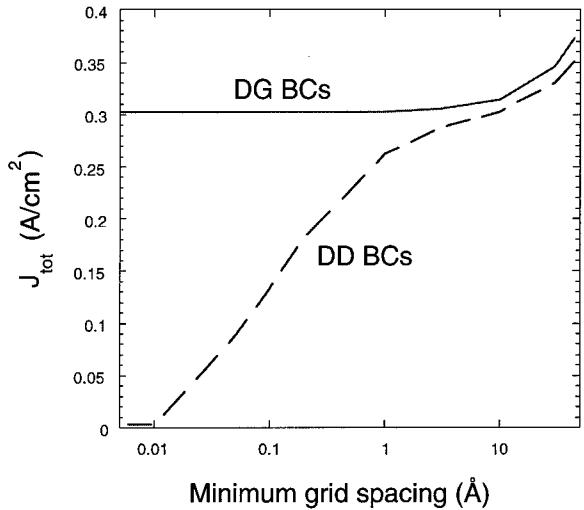


Figure 2. The dependence on grid spacing of the DG-simulated currents through a short pn diode. Note the large errors made if the DD BCs are used especially when the minimum grid spacing is very small.

(Fig. 1(b)) instead we find that the conventional ohmic conditions give grid-dependent results, a clear sign of their ill-posedness. The finer the mesh the narrower the boundary layer over which the solution adjusts to meet the inconsistent boundary condition and the worse the solution gets. The new DG BCs are instead seen to be well-behaved (i.e., give grid-independent solutions) yielding the profile also shown in Fig. 1(b).

The crucial importance of the minority carrier condition for the short diode is shown in Fig. 2 where we compare the current densities at $V_0 = 0.5$ V (forward bias) for a number of different minimum grid spacings (all grids are non-uniform with the finest grid spacing occurring at the contact). The poor performance of the conventional ohmic BCs is evident. Only at very large grid spacings do these conditions start to give reasonable results; in this case, discretization error is dominating and the new DG BCs do not perform as well because they have been discretized using a first-order formula (in contrast with the PDEs that are treated to second-order).

When the minority carrier diffusion length is decreased, bulk generation/recombination processes start to equilibrate the minority carriers and the inconsistency of the DD conditions (2b) has less impact because (i) the minority carrier density gradient is smaller so that the DG corrections are reduced and (ii) the contribution of minority carriers to the total current (at the contact) is smaller and hence errors in their

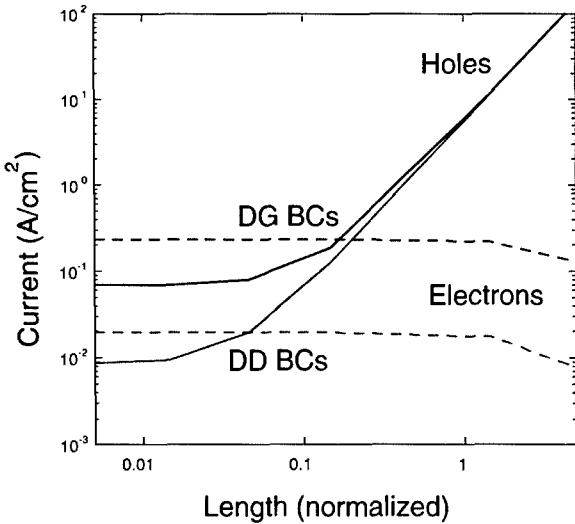


Figure 3. The effect of device size normalized by the diffusion length on the errors in the currents in a pn diode (at the contact to the *p*-region). When the device is large the majority carriers dominate and the error in the minority carrier current computed using the DD BCs, though still present, is irrelevant.

treatment are less relevant. To see these effects we plot (Fig. 3) the current density for a pn diode (at the contact to the *p*-region) as a function of the normalized thickness of the diode ($L/\sqrt{D\tau}$). A small minimum grid spacing (0.01 \AA) was used so as to magnify the error associated with the inconsistent DD ohmic BCs. As seen in the figure, as the normalized length increases above about one the electron contribution, though still in error, is insignificant.

A final illustration (Fig. 4) compares the errors in the I-V characteristic of the short diode (inset) as computed with the DD BCs (dashed) and with the DG BCs applied only to the minority carriers (solid). The pure DD case again shows the undesirable grid dependence previously noted (Fig. 1(b)). And when the DG BCs are applied solely to the minority carriers very little error is seen except at high forward voltages ($> 1.1 \text{ V}$). Since the latter condition produces high-level injection it seems reasonable that the DD BC (2b) applied to the majority carriers should produce errors because the particular value of N_A should no longer be relevant and the contact should instead act simply to source/sink carriers so as to preserve quasi-charge neutrality.

4. Insulator Proximity Effect

A second issue relating to ohmic contacts arises in 2-D/3-D problems in which the ohmic contact adjoins an

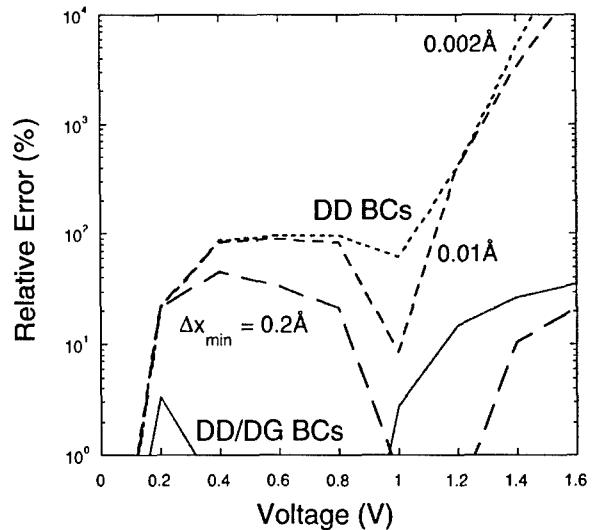


Figure 4. The relative errors in the I-V characteristics when either the DD BCs are used (dashed) or when the DD BCs are applied solely to the minority carriers (solid).

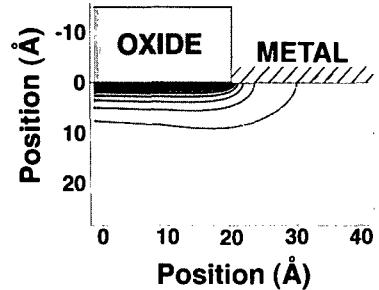


Figure 5. Electron density contours (maximum of $1.8 \times 10^{20} \text{ cm}^{-3}$ and spaced by factors of 1.8) near the corner where a MOSFET contact abuts the gate insulator calculated using the new DG ohmic BCs. Note the lateral repulsion of the carriers by the insulator at the contact edge.

insulator (see Fig. 5) so that there exists a “triple point” (or “triple edge”) where a semiconductor, a metal and an insulator all come together. To understand this issue, first note that in DG theory the BCs applied at a semiconductor-insulator junction are electrostatic conditions plus

$$\mathbf{n} \cdot \nabla \Phi_n = \mathbf{n} \cdot \nabla \Phi_p = 0, \quad n = p = n_{\text{small}} \quad (5)$$

where the first two conditions are zero current conditions (strictly valid only if no interface states are present) and the latter two, with n_{small} being a known small concentration (say, 1 cm^{-3}), approximate the

effect of barrier repulsion.² One can readily see that the ohmic BC on majority carriers used by DD theory ($p = p_{eq} \cong N_A$) would, at the edge of the ohmic contact, conflict with the insulator condition $p = n_{small}$. The discontinuity in p implies an infinite gradient and hence numerical problems. This trouble originates in the incorrect assumption that the ohmic contact is ideal right up to its edge. In fact, there will be an “insulator proximity effect” in which the insulator will repel the electrons in the adjacent metal thereby modifying the properties of the contact. (A similar repulsion of course occurs in the semiconductor, however, this is already treated by the DG equations (1)). A proper treatment of the situation therefore requires that the metal (and perhaps the insulator²) be treated as non-ideal. This can indeed be done within DG theory (Ancona 1992), but in keeping with the aforementioned crudeness of ohmic BCs we look for a way of modifying the BCs so as to retain an electrostatically ideal metal.

To this end, we note that using the new Neumann BCs in (4) at the metal interface eliminates the conflict noted in the previous paragraph with the Dirichlet conditions in (5) applied at the insulator interface. There is similarly no conflict between the conditions on Φ_n and Φ_p in (2a) and (5). So the only potential source of trouble is the condition on ψ . As discussed in Section 2, at an ohmic contact we take $\psi = V - \phi_{p0}(p)$ (where holes are the majority carrier). If this BC were to be used up to an insulator edge, the decrease of p from N_A to n_{small} caused by the insulator proximity effect would imply a variation in ψ which would violate the electrostatic BC (for an ideal metal) that the tangential component of the electric field should vanish. This inconsistency results because the DG correction in the condition on ψ (implied by the PDE (1b)₂) has been ignored. Including this term would, however, require treating the metal as non-ideal, so we instead eliminate the dependence on density entirely using the simple expedient:

$$\psi = \psi_c \cong V - \phi_{p0}(N_A) \quad (6)$$

where ψ_c is the value of ψ at the ohmic contact *far from the insulator*. The first equality represents a cumbersome non-local condition and so is best replaced (at least outside of the high-level injection regime

discussed earlier) by the simple approximation of the second equality. Obviously (6) will become problematic when the contact is so small that no part of it is far from the insulator, e.g., with a quantum point contact. In this case, it seems impossible to avoid treating the non-idealities of the metal (and perhaps the insulator).

To illustrate the DG treatment of a “triple point” numerically, in Fig. 5 we show a contour plot of the electron density near the corner where a MOSFET contact abuts the gate insulator. The calculations were performed using the simulation code PROPHET (Rafferty *et al.* 1998). The new DG ohmic BCs are seen to perform quite well, including exhibiting the carrier repulsion in the contact region associated with the insulator proximity effect.

Acknowledgment

The first author thanks the Office of Naval Research for funding support.

Notes

1. As usual, the condition on ψ derivable from the integral form of (1c) merely gives an equation with which one can determine the surface charge density at the contact *a posteriori*.
2. A proper treatment of barrier repulsion would include a non-ideal insulator that permitted barrier penetration (and could be treated using DG theory (Ancona 1990b)). However, for high barriers this would merely adjust the value of n_{small} , providing only a minor correction. In the case of an infinite insulator barrier, n_{small} vanishes.

References

- Ancona M.G. 1990a. Phys. Rev. B 42: 1222.
- Ancona M.G. 1990b. Superlatt. Microstruct. 7: 119.
- Ancona M.G. 1992. Phys. Rev. B 46: 4874.
- Ancona M.G. 2000. IEEE Trans. Elect. Dev. 47: 1449.
- Ancona M.G. and Iafrate G.J. 1989. Phys. Rev. B 39: 9536.
- Ancona M.G. and Tiersten H.F. 1987. Phys. Rev. B 35: 7959.
- Ancona M.G., Yu Z., Dutton R.W., Vande Voorde P.J., Cao M., and Vook D. 2000. IEEE Trans. Elect. Dev. 47: 2310.
- Asenov A., Slavcheva G., Brown A.R., Davies J.H., and Saini S. 2001. IEEE Trans. Elect. Dev. 48: 722.
- Rafferty C.S., Yu Z., Biegel B., Ancona M.G., Bude J., and Dutton R.W. 1998. In: Proc. SISPAD, Leuven, Belgium, p. 137.
- Wettstein A., Schenk A., and Fichtner W. 2001. IEEE Trans. Elect. Dev. 48: 279.



Molecular Devices Simulations Based on Density Functional Tight-Binding

ALDO DI CARLO, MARIETA GHEORGHE, ALESSANDRO BOLOGNESI AND PAOLO LUGLI
INFM-Dip. Ing. Elettronica, Università di Roma “Tor Vergata”, 00133 Roma, Italy

MICHAEL STERNBERG, GOTTHARD SEIFERT AND THOMAS FRAUENHEIM
Department of Physics, University of Paderborn, 33098 Paderborn, Germany

Abstract. We have developed a quantum simulation tool to investigate transport in molecular structures. The method is based on the joint use of a Density functional tight-binding (DFTB) and of a Green’s function technique which allows us the calculation of current flow through the investigated structures. Typical calculations are shown for carbon-nanotube-based field effect transistors, sensors and for DNA fragments.

Keywords: molecular electronics, tight-binding, transport

1. Introduction

Molecular electronics is attracting more and more attention both for its potential applications and for the interesting physical properties. Electronic conduction through a variety of different molecules has been studied experimentally by many groups. However the transport problem is still an open issue in these materials and a detailed microscopic investigation is necessary. In the following we will introduce a simulation approach able to describe the current flow in molecular structures. The approach is based on the density functional tight-binding description of the system coupled to a Green’s function technique.

2. Theory

The system we would like to describe can be generally divided in two parts: (i) the contacts a (ii) the molecular region. The contacts represent semi-infinite leads that end at the molecular region. On the other end, the molecular region can be any kind of atom collection such as the active part of a device or the molecule we would like to study via tunneling microscopy. In this work we consider the description of the system made via Density Functional Tight-Binding (DFTB) (Porezag *et al.* 1995, Elstner *et al.* 1998). We will not

enter in the details of the method which can be found in literature (Porezag *et al.* 1995, Elstner *et al.* 1998).

In order to solve the “current flow” problem, we need to use open boundary condition for the Kohn-Sham equations. Let’s consider the case with two contacts and a molecular region, under the assumption that there is no direct interaction between contacts. The hamiltonian for the full system can be described in blocks as follows

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_\alpha & \mathbf{T}_{\alpha M} & 0 \\ \mathbf{T}_{M\alpha} & \mathbf{H}_M & \mathbf{T}_{M\beta} \\ 0 & \mathbf{T}_{\beta M} & \mathbf{H}_\beta \end{bmatrix} \quad (1)$$

where $\mathbf{H}_{\alpha,\beta}$ the hamiltonian of the α, β contact, \mathbf{T} is the contact-molecule coupling Hamiltonian and \mathbf{S} is the overlap matrix

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_\alpha & \mathbf{S}_{\alpha M} & 0 \\ \mathbf{S}_{M\alpha} & \mathbf{S}_M & \mathbf{S}_{M\beta} \\ 0 & \mathbf{S}_{\beta M} & \mathbf{S}_\beta \end{bmatrix} \quad (2)$$

Now, from the equation of the Green’s function \mathbf{G} of the full system

$$\mathbf{G}^R = [(E + i\eta)\mathbf{S} - \mathbf{H}]^{-1} \rightarrow [(E + i\eta)\mathbf{S} - \mathbf{H}]\mathbf{G}^R = I \quad (3)$$

and defining the Self-Energy operator

$$\Sigma_{\alpha}^{R,A} = (E\mathbf{S}_{M\alpha} - \mathbf{T}_{M\alpha})\mathbf{g}_{\alpha}^{R,A}(E\mathbf{S}_{\alpha M} - \mathbf{T}_{\alpha M}) \quad (4)$$

we can express the Green's function of the molecular region as

$$\mathbf{G}_M^R = [E\mathbf{S}_M - \mathbf{H}_M - \Sigma^R]^{-1} \quad (5)$$

where

$$\Sigma^R = \sum_{\alpha} \Sigma_{\alpha}^R \quad (6)$$

Here \mathbf{g} is the the Green function of the uncoupled lead (see Di Carlo *et al.* (in press), Guinea *et al.* (1983), Lopez Sancho, Lopez Sancho and Rubio (1984, 1985)). As shown by the Eqs. (4) and (5) the contacts induce modification of the molecular region Green's function via self-energy terms. Such self-energy terms will only depend on the surface Green's function of the contact region. This follows from the nearest-neighbor interaction we have between the molecular region and the contacts (see Eq. (4)).

By using the defined Green function we can calculate the transmission coefficient between the α contact and the β contact (Datta 1995)

$$T_{\alpha\beta} = \text{Tr}[\Gamma_{\alpha}\mathbf{G}_M^R\Gamma_{\beta}\mathbf{G}_M^A] \quad (7)$$

where

$$\Gamma_{\alpha} = i[\Sigma_{\alpha}^R - \Sigma_{\alpha}^A] \quad (8)$$

From the knowledge of the transmission coefficient the coherent contribution to the current can be easily calculated via scattering theory. However, the advantage of Green function method consist in the possibility to extend the approach to treat non-coherent transport. This is accounted introducing the non-equilibrium Green functions (Datta 1995). Here we will not consider non-coherent transport (see Di Carlo *et al.* (in press) for explicit expression of non-equilibrium Green functions applied to DFTB).

In order to calculate the current flowing in the organic structure we should applied an external bias. This can be introduced in an approximate form which has been discussed by Datta and coworker (1997) or, as we did, fully accounted in the Hamiltonian:

$$H_{ij}^{(V)} = H_{ij} + \frac{e}{2}(V_i + V_j)S_{ij} \quad (9)$$

3. Applications

In the following we will present some applications of the theory we have discussed in the last section.

As a first example we show the calculation of the current in a nanotube-based Field Effect Transistor (FET). The structure of the simulated devices is shown in the inset of Fig. 1 where we use a fluorinated nanotube (Seifert, Köhler and Frauenheim 2000) to connect the source and drain contacts. As in conventional FET the current is modulated by the gate electrode. Similar devices, with carbon nanotubes, has been investigated experimentally by several authors (Tans *et al.* 1997, Bockrath *et al.* 1997, Tans, Verschueren and Dekker 1998, Martal *et al.* 1998). The calculated drain-source current for two drain bias as a function of the gate bias is shown in Fig. 1. For a given drain bias the device presents two well distinct regions, with the current saturating for negative gate bias and being reduced for positive gate bias. For a $V_{DS} = 2$ V the current is essentially negligible and we can consider that the "channel" of the FET is pinched-off. By reducing the drain bias the current also diminishes. Indeed, we observe an almost linear dependence of the I_{DS} in the saturation region as function of V_{DS} . We should point out that the results shown are in good agreement with those reported in the literature (Martel *et al.* 1998).

As a second example of calculation we show the calculated current along a DNA fragment. The structure is reported in Fig. 2. Here we consider a fragment of DNA with a single Guanine. A sulphur atom has been added at the beginning and at the end of the fragment in order to bound the gold contacts (Tian *et al.* 1998, Pantelides, Di Ventra and Lang 2001). In this calculation we did not

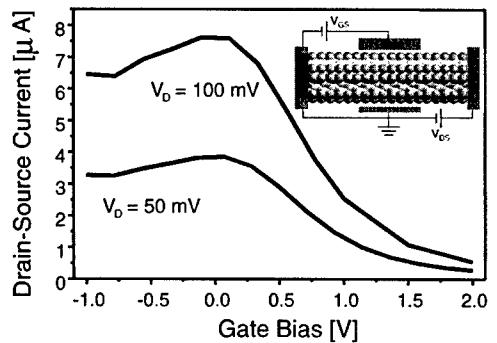


Figure 1. Calculated drain current calculation as a function of the gate bias for two drain bias in the FCN-FET. Inset: Schematic drawing of a fluorinated carbon nanotube based field effect transistor (FCN-FET).

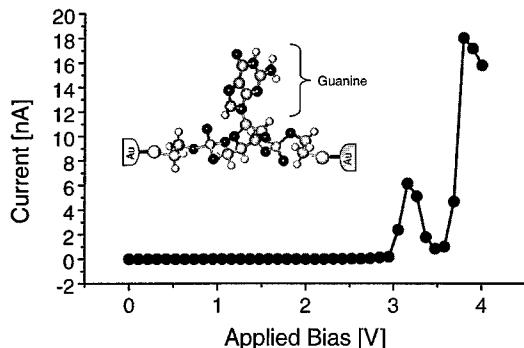


Figure 2. Calculated current in a DNA fragment. The DNA fragment structure is shown in the inset.

account for the gold atoms which are treated within an s-band approximation (Tian *et al.* 1998, Pantelides, Di Ventra and Lang 2001). The calculated current (Fig. 2) shows a well pronounced peak around 3 V which marks a resonant tunneling through the molecular orbitals.

In the TB picture the total energy can be written as the sum of the band structure energy and the repulsive energy U (Menon and Allen 1986, Sankey and Allen 1986, Tomanek and Schlüter 1987, Majewski and Vogl 1989)

$$E = \sum_k n_k \epsilon_k + \sum_{l>l'} U(R_{ll'}) \quad (10)$$

where n_k is the occupancy of the electronic state labeled by k and $R_{ll'}$ is the separation of ions l and l' . With this expression of the total energy and by using the Helmann-Feynman theorem one can calculate the force acting on the j -th ion

$$M\ddot{\mathbf{R}}^{(j)} = - \sum_k n_k \left\langle \Psi_k \left| \frac{dH}{d\mathbf{R}^{(j)}} \right| \Psi_k \right\rangle - \frac{dU}{d\mathbf{R}^{(j)}} \quad (11)$$

From the knowledge of forces, Molecular Dynamic (MD) simulations may be easily performed by using special integration algorithms (like the Verlet algorithm) (Turchi, Gonis and Colombo 1998, 2000). Beside typical applications of molecular dynamic simulations (Turchi, Gonis and Colombo 1998, 2000), MD can be coupled with current calculations. Kubo-Greenwood equation has been recently used (Kaschner *et al.* 1996, Seifert *et al.* 1998) to evaluate electric conductivity over MD trajectory in liquid NaSn alloys. This allows to evaluate the thermal average of the conductivity as a function of the lattice temperature. We also have applied this method to evaluate thermal averaged currents in carbon nanotubes (CNT).

However in the following we will present the simulations of a time dependent current response of a (5,5) CNT which reacts with a C_6H_4 molecule (benzyne). This kind of interaction has been investigated by means of classical MD simulations by J. Han and coworkers (1997). In our simulation we have used the DFTB with a minimal basis. Current flowing in the CNT has been obtained by applying the Green Function technique in order to ensure proper boundary conditions (see Di Carlo *et al.* 2002). The nanotube is biased with a linear potential drop of $V = 100$ meV along the tube direction. In these simulations the left and right contacts of the nanotube (two unit cells of the nanotube on each side) are kept fixed (carbons do not move) while the middle part of the CNT (the one which reacts with the C_6H_4 molecule) is free to move. The benzyne molecule moves towards the CNT with an initial velocity of $0.6 \text{ \AA}/\text{ps}$ as shown in the inset of Fig. 3.

During the MD evolution the current is calculated each 0.9675 fs. In order to have an idea about the size of the system, here we consider about 130 atoms with 508 orbitals all together. The calculation is performed with a standard workstation and 1 ps of simulation are obtained in few hours. Figure 3 shows the calculated current as a function of time. As soon as the molecule reaches the nanotube a strong reduction of the current flowing along the CNT occurs. After molecular rearrangement, which includes also a rotation of the C_6H_4 molecule, the system reaches its stationary condition after 0.2 ps. In Fig. 3 we also show the calculated current flowing in the CNT without the benzyne for a temperature of 300 K. Comparing the calculated current with and without the benzyne molecule we can see

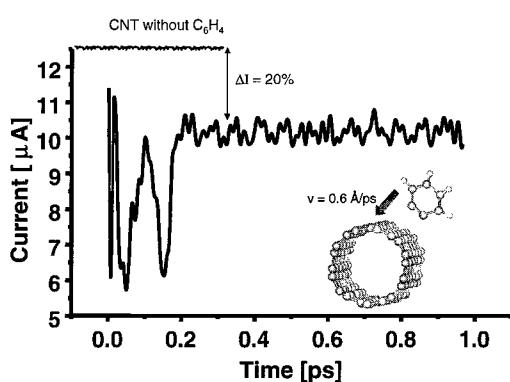


Figure 3. Calculated current flow in the biased ($V = 100$ meV) CNT + C_6H_4 molecule as a function of time. The upper curve represents the room temperature current in the biased CNT without the C_6H_4 molecule.

a strong reduction of the current flowing in the CNT when benzyne is attached to the CNT. In this case we can say that the benzyne on a CNT behaves as a defect that increases the resistance of the CNT (there is a variation of 20% in the resistance). Such effect is much more pronounced in the initial stage of the reaction where a local strong distortion of the CNT is also observed. Moreover, the structures of the current as a function of time also reflect the relaxation process that the absorbed molecule is undertaking. This simulation shows a possible application of CNT as gas sensor which is obtained by monitoring the current flowing in the CNT.

In conclusion we have shown that density functional tight binding methods can be efficiently coupled to Green's function techniques to account for current flow in organic nanostructures. We have shown typical application to organic and biological structures and in particular we have discussed the gate action in a nanotube based FET.

Acknowledgment

This work partially supported by the Office of Naval Research (ONR), the European DIODE Network and by the Italian MURST-CNR under the project "Nanoelectronics".

References

- Bockrath M. *et al.* 1997. *Science* 275: 1922.
- Datta S. 1995. *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, Cambridge, UK, Chap. 3 and 8.
- Cambridge Studies in Semiconductor Physics and Microelectronic Engineering, Vol. 3.
- Datta S. *et al.* 1997. *Phys. Rev. Lett.* 79: 2530.
- Di Carlo A., Gheorghe M., Lugli P., Sternberg M., Seifert G., and Frauenheim T. 2002. *Physica B* 314: 86.
- Elstner M. *et al.* 1998. *Phys. Rev. B* 58: 7260.
- Guinea F., Tejedor C., Flores F., and Louis E. 1983. *Phys. Rev. B* 28(8): 4397.
- Han J., Globus A.I., Jaffe R., and Deardorff G. 1997. *Nanotechnology* 8: 95. Also at www.nas.nasa.gov/Groups/Nanotechnology/publications/MGMS_EC1/simulation/data/
- Kaschner R., Schöne M., Seifert G., and Pastore G. 1996. *J. Phys.: Condens. Matter* 8: 653.
- Lopez Sancho M.P., Lopez Sancho J.M., and Rubio J. 1984. *J. Phys. F: Metal Physics* 14(5): 1205.
- Lopez Sancho M.P., Lopez Sancho J.M., and Rubio J. 1985. *J. Phys. F: Metal Physics* 15(4): 851.
- Majewski J.A. and Vogl P. 1989. In: de Boer F.R. and Pettifor D.G. (Eds.), *The Structure of Binary Compounds*. Elsevier, Amsterdam.
- Martel R. *et al.* 1998. *App. Phys. Lett.* 73: 2447.
- Menon M. and Allen R.E. 1986. *Phys. Rev. B* 33: 7099.
- Pantelides S.T., Di Ventra M., and Lang N.D. 2001. *Physica B* 296: 72.
- Porczag D. *et al.* 1995. *Phys. Rev. B* 51: 12947.
- Sankey O.F. and Allen R.E. 1986. *Phys. Rev. B* 33: 7164.
- Seifert G., Kaschner R., Schöne M., and Pastore G. 1998. *J. Phys.: Condens. Matter* 10: 1175.
- Seifert G., Köhler T., and Frauenheim T. 2000. *App. Phys. Lett.* 77: 1313.
- Tans S.J. *et al.* 1997. *Nature* 386: 474.
- Tans S.J., Verschueren A.R.M., and Dekker C. 1998. *Nature* 396: 49.
- Tian W. *et al.* 1998. *J. Chem. Phys.* 109: 2874.
- Tomanek D. and Schlüter M.A. 1987. *Phys. Rev. B* 36: 1208.
- Turchi P.E.A., Gonis A., and Colombo L. (Eds.). 1998. *Tight-Binding Approach to Computational Material Science*. MRS Proc. 491.



Role of Carrier Capture in Microscopic Simulation of Multi-Quantum-Well Semiconductor Laser Diodes

M.S. HYBERTSEN*

Agere Systems, 600 Mountain Ave., Murray Hill, NJ 07974, USA

msh@agere.com

B. WITZIGMANN

Agere Systems, Alhambra, CA 91803, USA

M.A. ALAM AND R.K. SMITH

Agere Systems, 600 Mountain Ave., Murray Hill, NJ 07974, USA

Abstract. Microscopic laser simulation for multi-quantum-well devices must include conventional carrier transport, properties of carriers bound in the quantum wells, and the photon modes in the optical cavity. Physical models unique to the laser simulation problem include capture of carriers into the quantum wells and stimulated emission, both fundamentally requiring quantum mechanical calculations. The implementation of a semiclassical carrier capture model into a fully self-consistent laser simulator is discussed. The impact of the capture process is illustrated for a two-quantum-well laser.

Keywords: semiconductor laser simulation, quantum-well laser, carrier capture

1. Introduction

The substantial challenges in design and manufacture of laser diodes for telecommunications applications motivate continued research to better understand the fundamentals of laser diode operation. Microscopic laser simulation wherein one attempts to integrate all of the relevant physical models into a self-consistent treatment of the laser diode operation has been actively pursued by a several research groups in recent years (Li *et al.* 1992, Tessler and Eisenstein 1993, Grupen and Hess 1998, Alam *et al.* 2000, Witzigmann, Witzig and Fichtner 2000, Piprek, Abraham and Bowers 2000). The case of laser devices with quantum well regions in the active layer presents special challenges for simulation. Not only are the length scales short (order 10 Å), but one must integrate fundamentally quantum

mechanical processes (carrier capture into the quantum wells and stimulated emission) into the simulation that is otherwise based on a semiclassical formulation. Nonetheless, it was recognized early in the development of quantum well lasers that the time scale for carrier capture into the quantum well could be significant to the device performance. Rate equation analysis suggested that the modulation response of the device could be severely impaired (Nagarajan *et al.* 1992).

In this paper, we briefly review the essential features of our *LASER* simulator (Alam *et al.* 2000). The simulator has been extensively applied to study practical multi-quantum-well (MQW) laser performance. The models in the simulator have been verified by comparison to extensive data (DC terminal characteristics, optical characteristics and AC characteristics) as a function of internal parameters (doping in the active region) and temperature (Hybertsen *et al.* 1999, Witzigmann *et al.* unpublished). We then describe in more detail the

*To whom correspondence should be addressed.

implementation of a semiclassical carrier capture model, including some of the numerical difficulties involved. The physical impact of the capture model is illustrated for a two-quantum-well laser example.

2. Overview of LASER

The bulk carriers are treated with a conventional drift-diffusion formulation. In the regions with quantum wells, there are additional bound carrier populations introduced (electrons and holes). The bulk carriers move over a continuation of the barrier band edge, but in addition experience a generation-recombination process that couples the bulk carriers to the bound carrier populations. This model is described in detail below. The properties of the bound carriers are calculated quantum mechanically using an eight band $k \cdot p$ model. The electrostatic potential in the region of the quantum well is included selfconsistently: the potential is included in the Hamiltonian for the bound carriers and the bound carrier distributions are included in the Poisson equation. The electronic states are used to determine the 2D quasi-Fermi levels for the bound populations and to compute the optical gain and radiative recombination. Both bulk and bound carriers also undergo non-radiative recombination. The transport simulation accounts for processes in the cross-section of the laser diode. In a typical semiconductor laser diode design, only one or a few optical modes are guided so as to have substantial optical gain. The distribution of the optical modes is calculated by solution of the Helmholtz equation, accounting for the indices of refraction (carrier density dependent) in the various regions of the device cross-section. For the edge emitting laser diodes studied here, the resonant cavity is formed orthogonal to this cross-section. The optical modes formed by this cavity are treated as discrete resonances with an associated cavity loss. The population of these modes is calculated from rate equations that include the stimulated emission from the quantum well regions. More details may be found in Alam *et al.* (2000).

The physical description of the laser has been reduced to a set of coupled integro-differential equations that must be solved selfconsistently. The equations are solved iteratively. The eigenvalue equations (Helmholtz equation for the optical mode and Schrödinger equation for each quantum well) are solved first to give input to the physical models. Then a Newton update step is performed on the drift-diffusion equations and the rate equations for the quantum well populations

and the photon modes. The solution variables $u = (n_{3D}, p_{3D}, V, n_{2D}, p_{2D}, S)$ (bulk electron density, bulk hole density, electrostatic potential, bound electron density for each well, bound hole density for each well and photon number for each mode) are organized so that the necessary Jacobian can be factored in block form:

$$g'(u) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (1)$$

Here the A_{11} block is the Jacobian of the drift-diffusion equations with the usual sparse data structure. The A_{22} block spans the rate equations for the bound carriers and the photon modes. This block and the coupling terms (A_{12} and A_{21}) are dense. However, since the size of the dense block is relatively small, the Jacobian can be efficiently managed. The pole in the photon rate equation can pose special difficulties for an iterative solution method. These problems are overcome by introducing a slack variable for each photon mode and augmenting the equation set. Further details are given in Alam *et al.* (1997, 2000). The implementation of the capture model is described below.

Most of the parameters that enter the physical models are taken from the literature (band masses, mobilities, recombination coefficients, etc.) However, a few key parameters are either not known with sufficient accuracy from existing independent experiments or are unique to our formulation of the model. These parameters are fixed by reference to selected device measurements as follows. The coefficient for free-carrier absorption in the quantum wells (dominated by intervalence band absorption) is determined from direct measurements of cavity loss versus current below threshold. The Auger recombination coefficients are fixed by the threshold current density of wide area devices. The temperature dependence of the Auger recombination is determined by a single activation energy. With this choice, the temperature dependence of the threshold current density is accurately simulated. A gain compression coefficient is included to account for carrier heating and spectral hole burning in the bound carrier populations, processes that are not explicitly included in the simulator. The magnitude is fixed by the scale of the damping of the modulation response of the laser. Finally, the scattering time parameter (τ in Eq. (3) below) for the carrier capture model must be chosen. It is hard to find a direct, easily interpreted measurement to fix this rate. However, various device characteristics can be used to constrain the values (Hybertsen *et al.* 2000).

We have shown that having fixed these parameters for a particular $1.3 \mu\text{m}$ MQW laser design, the simulator accurately predicts the performance of a whole family of these devices in which the doping in the active layer and the temperature were systematically varied (Hybertsen *et al.* 1999, Witzigmann *et al.* unpublished).

3. Carrier Capture Model

A key issue in formulating the capture of carriers into quantum wells is the coherence length. It is common to treat the bulk carriers as coherent over a long range. This can lead to unphysical dependence of the results on boundary conditions far from the quantum well. There has only been limited work to self consistently include the dephasing intrinsic to the scattering process by which a carrier is captured (Register and Hess 1997). In a forward biased laser, there is the further complication of a high carrier density in the region of the quantum well. In the context of device simulation, where a semiclassical formulation is already used for the bulk carriers, it is natural to assume sufficiently rapid scattering so that the coherence length is short, even compared to the quantum well dimension.

In the laser, both carrier-phonon and carrier-carrier scattering play a role in the carrier capture. In the Mini-Lase simulator (Grupen and Hess 1998), matrix elements for both processes are estimated and the net scattering into the bound population is resolved into energy bins. Furthermore, scattering between different energy bins is also considered. This has allowed the simulator to resolve the delicate interplay between carrier heating (cooling) and spectral hole burning (Grupen and Hess 1998).

We have adopted a more phenomenological approach. The function $S(E_2, E_3)$ describes the probability of scattering from the bulk state at energy E_3 to the bound state at energy E_2 . Then, including the densities of states and the Fermi functions, the net capture rate can be written as

$$C = \int_{E_b}^{\infty} dE_3 \int_{E_w}^{\infty} dE_2 g_2(E_2) g_3(E_3) S(E_2, E_3) \times f_3(E_3)(1 - f_2(E_2)). \quad (2)$$

In general, there are both elastic and inelastic processes contributing to the capture process. The expression can be simplified by restricting S to describe elastic scattering. An energy dependence in the scattering rate can

be included to account for reflection at the well and phase-space reduction (Baraff 1997, 1998):

$$g_2(E_3) g_3(E_3) S(E_3) \propto E_3^{\beta+1/2}. \quad (3)$$

Here $\beta = 1$ accounts for both. Finally, the matrix element is parameterized by a single time constant leading to:

$$C = (F_{3/2}(\eta_{3D}) - F_{3/2}(\eta_{2D})) / \tau \quad (4)$$

where $\eta_{2D(3D)} = (F_{2D(3D)} - E_b) / kT$ and the Fermi function of order 3/2 enters for $\beta = 1$. To this point, we have not considered the coherence volume. In practice, we assume the overall dephasing time in the region near the quantum well to be sufficiently short, that we can interpret Eq. (4) to apply locally (Baraff 1997). In this case, the quasi-Fermi level F_{3D} and the barrier band edge E_b become a function of position, as they normally are considered in the drift-diffusion formulation of transport. This situation is illustrated in Fig. 1, showing the local effect of the electrostatic potential and a non-uniform quasi-Fermi level for the bulk carriers (F_{3D}). In this case, it is also consistent to return to Eq. (3) and interpret the density of states factors as local. The bound carriers are extended according to the subband wavefunctions:

$$g_2(E_2, y) \propto |\phi(y)|^2 \theta(E_2 - E_{2D}). \quad (5)$$

The position dependence from Eq. (5) flows through to Eq. (4) contributing a locally position dependent prefactor. The final result is intuitively reasonable: the bulk carriers experience a local generation-recombination

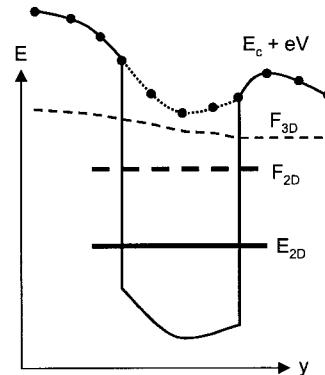


Figure 1. Schematic diagram illustrating the band diagram near a quantum well, the bound level and the associated quasi-Fermi levels.

process that depends on the quasi-Fermi level separation relative to the bound carriers and the probability density for the bound carriers at that point.

Implementation of Eq. (4) presents some complications. The capture process $C(V, F_{3D}, F_{2D})$ must be recast in terms of solution variables for the Newton update scheme described in Section 2: V , n_{3D} and n_{2D} . When this is done, one sees that F_{2D} depends on the electrostatic potential V on the entire set of grid points spanning the region that encompasses the quantum well. Therefore, when evaluating Eq. (4) on a particular node, it will depend on the electrostatic potential V on many neighboring nodes (Fig. 1). This degree of non-locality goes well beyond that normally required in a drift-diffusion formulation. The sparse structure for the drift-diffusion part of the Jacobian (A_{11}) in Eq. (1) would be significantly altered. In order to avoid the extra nonlocality in the capture model, we introduce auxiliary solution variables F_{2D} for electrons and holes in each well region. With this augmented set of variables, we restore the sparse data structure for the bulk solution variables (A_{11}) at the cost of a negligible increase in the dense matrix size (A_{22}).

4. Two-Quantum-Well Example

We illustrate the impact of the carrier capture model using a simplified example with two quantum wells in the active region. The quantum wells are typical of a $1.3\text{ }\mu\text{m}$ InP based laser. The active layer includes separate confinement layers (1000 \AA each) of the same composition as the barrier material. To achieve threshold at reasonable carrier density levels, we choose a small mirror loss (10 cm^{-1}) and neglect free carrier absorption.

We start by considering the regime of slow carrier capture, choosing the scattering time parameter in the capture model (τ) to be 10 ps. Figure 2 shows the band diagram in the active region at a current substantially above threshold. In this example, the active layer is undoped. Due to the relatively slow capture coefficient, the quasi-Fermi level separation is forced to be larger to support the current flow into the quantum wells required for the stimulated emission. With a slow capture rate, a substantial fraction of the carriers pass over the well. The bound carriers are evenly divided between the wells and the gain contributed is about the same from each well. Furthermore, the carriers flow past the wells and fill the separate confinement layers. The shallow electron wells also contribute to the

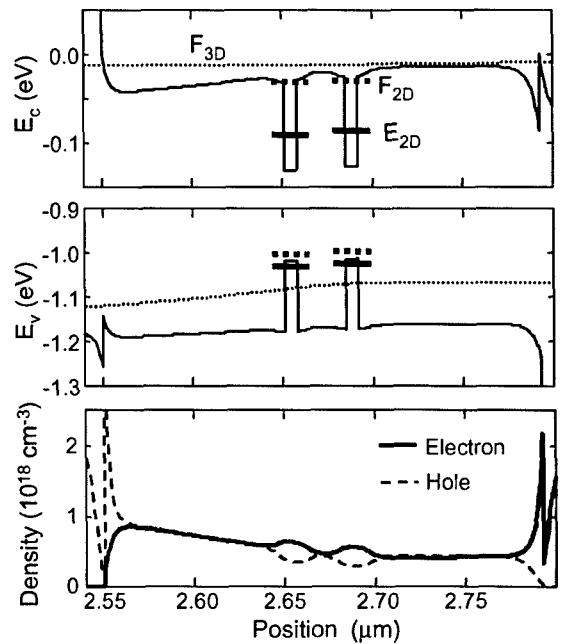


Figure 2. Band diagram and bulk carrier densities in the active region of a two-quantum-well laser. The scattering time parameter in the capture model (τ) was chosen to be 10 ps.

large bulk carrier population. Note that the bound electron quasi-Fermi levels are near the tops of the wells. This carrier spill out contributes to the local electrostatic potential and attracts extra bulk electrons to the regions over the wells (bottom of Fig. 2). The bulk carrier population rises with current, so there is a substantial capacitance associated to this. This capacitance can be largely eliminated by doping the separate confinement layers. Overall, with a slow capture process, the transport in the active region can be viewed as a large bulk population above the wells that feed the bound carrier populations approximately equally through the capture process.

A different qualitative picture emerges for the fast capture limit. We choose the scattering time parameter in the capture model (τ) to be 0.1 ps. In this simulation, the separate confinement layers are doped. Figure 3 shows the quantum well region for the same bias as in Fig. 2. The 2D quasi-Fermi levels are pinned very close to the 3D quasi-Fermi level. In the hole case, the 3D quasi-Fermi level crosses the 2D quasi-Fermi level for the left quantum well. Physically, there is capture on the left side of the well and reemission on the right side. This is intuitively reasonable. With fast capture, the hole transport in particular will be sequential. As a

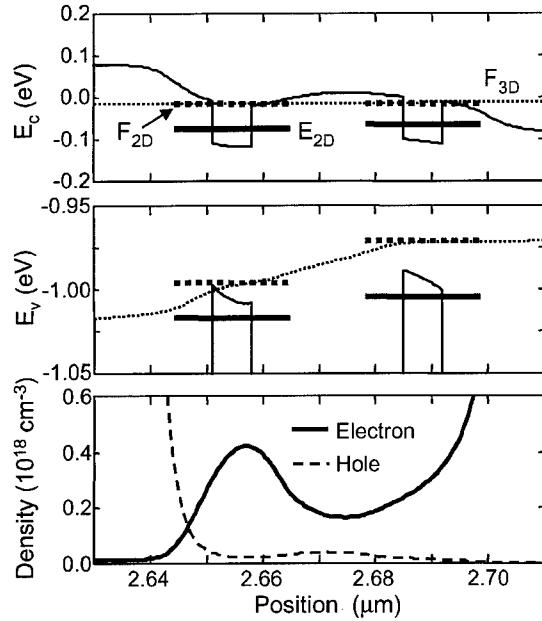


Figure 3. Zoom in on the band diagram and bulk carrier densities near the wells for a two-quantum-well laser. The scattering time parameter in the capture model (τ) was chosen to be 0.1 ps. Doping was extended through the active to within 3 nm of the edge of the well regions as is apparent from the rise in the free carrier densities at edges of the region shown.

consequence, the bound carrier population is larger in the left quantum well and most of the gain comes from the left well (80%). The local electrostatic potential and bulk carrier densities also reflect the larger bound hole population in the left well.

The different regimes of carrier capture illustrated here yield quite different modulation response. Figure 4 shows a comparison of the small signal response for identical current bias conditions. Both devices have doped separate confinement layers to minimize any impact from the bulk carrier accumulation described in Fig. 2. In the slow capture case, the capture process is the gating process in the modulation response. This leads to the heavy low frequency roll-off that largely suppresses the resonance. This is most apparent in the phase of the response that shows a large linear component and almost no phase shift at the resonance (near 10 GHz). In the fast capture case, the amplitude modulation shows the classic resonance. The impact of the transport can only be qualitatively discerned from the phase that shows a linear component. The phase shift at resonance (also near 10 GHz) is substantially less than 180 degrees.

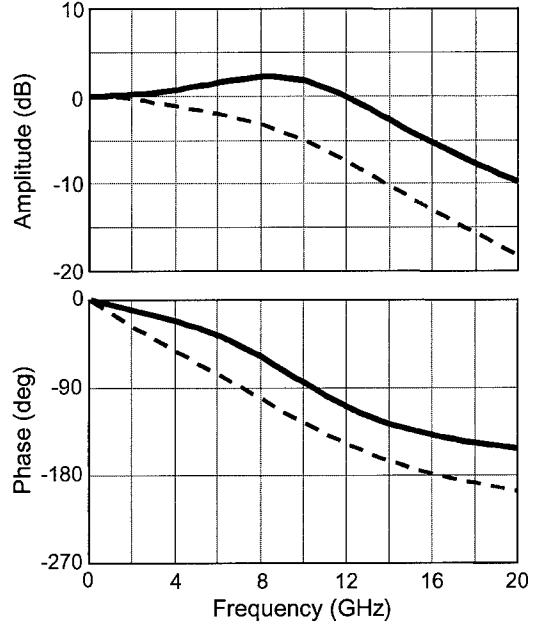


Figure 4. Modulation response (amplitude and phase) simulated for scattering time parameter (τ) in the capture model of 10 ps (dashed lines) and 0.1 ps (solid lines).

5. Conclusions

We have discussed the implementation of a semiclassical carrier capture model into a full microscopic laser simulator. A tutorial example illustrates different regimes of laser operation depending on the time scale for the capture process. In practice, we have found a scattering time parameter (τ) intermediate to those discussed here (1 ps) to be appropriate to the description of InGaAsP based MQW devices (Alam *et al.* 2000, Hybertsen *et al.* 1999, 2000, Witzigman *et al.* unpublished). The tutorial example suggests that the extended set of rate equations commonly employed to include the carrier capture process phenomenologically (Nagarajan *et al.* 1992, Esquivias *et al.* 1999) may not be the best choice for a compact model of MQW laser modulation. A microscopic laser simulator, such as the one described here, can be used as a tool to investigate compact models that are more representative of the dynamic response of telecommunications wavelength lasers.

References

- Alam M.A., Hybertsen M.S., Smith R.K., and Baraff G.A. 2000. IEEE Trans. on Electron Devices 47: 1917.
- Alam M.A., Hybertsen M.S., Smith R.K., Baraff G.A., and Pinto M.R. 1997. In: Osinski M. and Chow W. (Eds.), Physics and

- Simulation of Optoelectronic Devices V, Proceedings of SPIE. Vol. 2994, p. 709.
- Baraff G.A. 1997. Phys. Rev. B 55: 10745.
- Baraff G.A. 1998. Phys. Rev. B 58: 13799.
- Esquivias I., Weisser S., Romero B., Ralston J.D., and Rosenzweig J. 1999. IEEE J. of Quantum Electron. 35: 635.
- Grupen M. and Hess K. 1998. IEEE J. Quantum Electron 34: 120.
- Hybertsen M.S., Alam M.A., Baraff G.A., Smith R.K., Shtengel G.E., Reynolds C.L. Jr., and Belenkay G.L. 2000. In: Binder R.H., Blood P., and Osinski M. (Eds.), Physics and Simulation of Optoelectronic Devices VIII, Proceedings of SPIE. Vol. 3944, p. 486.
- Hybertsen M.S., Alam M.A., Shtengel G.E., Belenkay G.L., Reynolds C.L., Smith R.K., Baraff G.A., Kazarinov R.F., Wynn J.D., and Smith L.E. 1999. In: Blood P., Ishibashi A., and Osinski M. (Eds.), Physics and Simulation of Optoelectronic Devices VII. Proceedings of SPIE. Vol. 3625, p. 524.
- Li Z-M., Dzurko K.M., Delage A., and McAlister S.P. 1992. IEEE J. Quantum Electron. 28: 792.
- Nagarajan R., Ishikawa M., Fukushima T., Geels R.S., and Bowers J.E. 1992. IEEE J. Quantum Electron. 28: 1990.
- Piprek J., Abraham P., and Bowers J.E. 2000. IEEE J. of Quantum Electron. 36: 366.
- Register L.F. and Hess K. 1997. Appl. Phys. Lett. 71: 1222.
- Tessler N. and Eisenstein G. 1993. IEEE J. Quantum Electron. 29: 1586.
- Witzigmann B., Hybertsen M.S., Reynolds C.L., Belenkay G.L., Shterengas L., and Shtengel G.E. unpublished.
- Witzigmann B., Witzig A., and Fichtner W. 2000. IEEE Trans. on Electron Devices 47: 1926.



Numerical Study of Minority Carrier Induced Diffusion Capacitance in VCSELs Using Minilase

YANG LIU

*Beckman Institute and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA*

FABIANO OYAFUSO

High Performance Computing Group, Jet Propulsion Laboratory, Pasadena, CA 91109, USA

WEI-CHOON NG AND KARL HESS

*Beckman Institute and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA*

Abstract. Our numerical simulations of vertical cavity surface emitting lasers with Minilase demonstrate that a diffusion capacitance is induced by the minority carriers accumulated in the separate confinement regions. This diffusion capacitance is shown to be responsible for the over-damping of the modulation response and the reduction of the modulation bandwidth. It is also demonstrated that this diffusion capacitance is significantly suppressed by grading the separate confinement hetero-junctions (SCHs) or reducing the thickness of the SCHs.

Keywords: numerical simulations, vertical cavity surface emitting lasers, minilase, modulation response, diffusion capacitance, minority carriers

1. Introduction

Vertical cavity surface emitting lasers (VCSELs) have emerged recently as a promising light source for applications in optical communications. As the high-speed performance becomes one of VCSELs' key issues, numerical simulations of their modulation responses have become indispensable for both theoretical understanding and device design optimization. Although rate equation models are numerically tractable, a phenomenological nonlinear gain suppression factor has to be introduced to account for the damping in the actual response curves. Our comprehensive laser diode simulator, Minilase, has adopted a different approach and can simulate nonlinear gain and modulation response of laser diodes from very basic principles (Grupen and Hess 1998). Recently, Minilase has been extended to simulate the fully coupled electrical and optical systems in VCSELs (Oyafuso *et al.* 2000).

In this paper, we focus on the effect of the minority carriers on the modulation response of VCSELs. Some simulation works related to this effect have previously been done for edge-emitting laser diodes based on simple rate equation models (Nagarajan *et al.* 1992, Rideout *et al.* 1991). In Nagarajan *et al.* (1992), the carrier escape time τ_e and capture time τ_s have been introduced in a phenomenological way. As presented in Nagarajan *et al.* (1992), the analytical solution to these rate equations gives a reduced differential gain by a factor of $\chi = 1 + \tau_s/\tau_e$, resulting in the reduction of the relaxation frequency for the same power level. It is shown in Rideout *et al.* (1991) that the finite carrier capture rate from the barrier to the quantum well can induce a well-barrier hole burning effect and result in the roll-off of the response curves. However, these rate equation models are over-simplified since only one type of carrier is assumed and the minority carriers are not separated from the majority carriers. Furthermore, the

dependencies of τ_c on the device structure and other gain suppression mechanisms, such as spatial hole burning (SHB) effects in VCSELs, are not considered. In Minilase, the escape of minority carriers from the quantum well (QW) is treated as thermionic emission (Grupen and Hess 1998) and its dependence on the device structure and SHB effect are self-consistently included. The carrier flux of the thermionic emission from the QW to the barrier is expressed as (Hess 1988)

$$j_{w \rightarrow b} = A^* T^2 \left[\exp\left(\frac{F_w - E_w - \Delta E}{k_B T}\right) - \exp\left(\frac{F_b - E_b}{k_B T}\right) \right] \quad (1)$$

where ΔE is the band-edge discontinuity at the well/barrier interface, and $A^* = m^* k_B^2 / (2\pi^2 \hbar^3)$ is the Richardson constant. Our Minilase simulations with various VCSEL structures demonstrate that a diffusion capacitance is induced by the accumulation of wasted minority carriers in the SCHs. This diffusion capacitance is shown to be responsible for the reductions of both the modulation bandwidth and the height of the relaxation peak. Further simulations with Minilase also indicate that this damping effect is greatly dependent on the grading and the thickness of the SCHs.

2. Simulation Results

We base Minilase simulation on a double oxide-confined 1- λ VCSEL structure, which experimentally showed a bandwidth greater than 16 GHz (Lear *et al.* 1996). Two optical solvers have been developed, one based on a Green's function method (Klein *et al.* 1998), and the other on a scalar effective index method. In order to obtain the optical intensity pattern, the lasing frequency, and the optical loss, the entire device structure including the two distributed Bragg reflector (DBR) stacks is considered in our optical solvers. For the electronic part, since our major interest is the intrinsic dynamics of the laser diodes, we only simulate the cavity between the top and the bottom oxide layers, the schematic cross-section of which is shown in Fig. 1. The VCSEL structure has a single 80 Å In_{0.2}Ga_{0.8}As QW layer and operates at a wavelength of 980 nm. The oxide apertures are composed of Al_{0.6}Ga_{0.4}As and the radius is 1.8 μm. For comparison purposes, an artificial 2- λ cavity VCSEL is also simulated, the SCH

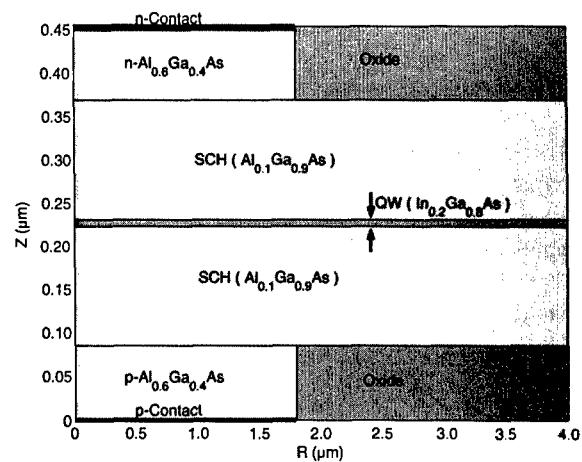


Figure 1. Schematic cross-section of a double oxide-confined 1- λ cavity VCSEL structure, which is simulated with the electronic solver of Minilase.

thickness of which is approximately twice that of the 1- λ cavity structure. For both cavities, either ungraded SCHs of homogeneous Al_{0.1}Ga_{0.9}As or SCHs linearly graded from Al_{0.1}Ga_{0.9}As to Al_{0.6}Ga_{0.4}As are used. To focus on the effects due to the electrical transport, we assume that the optical confinement factor, the optical transverse mode pattern within the QW and the optical loss remain unchanged for all these structures.

Figure 2 shows the simulated minority electron flux contours for the un-graded, 2- λ cavity VCSEL at a

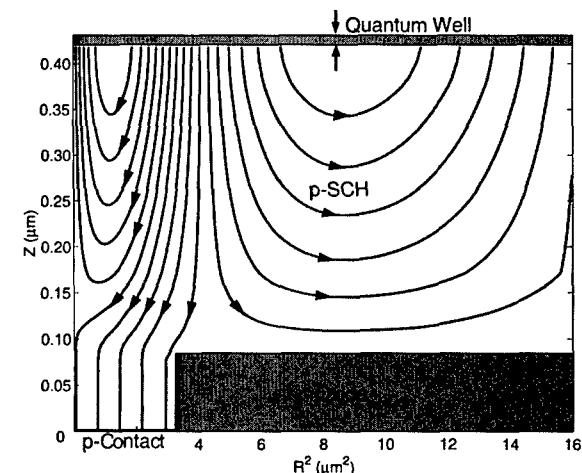


Figure 2. Contour plot of the simulated minority electron flux in the p-side separate confinement region for an un-graded, 2- λ cavity VCSEL operating at $2 \times I_{th}$. Note that we use the radius square as the x-axis to correctly account for the cylindrical geometry of VCSELs.

current bias twice the threshold current. It is worth noting that the square of the radius has to be used as the x -axis to correctly account for the cylindrical geometry of VCSELs while plotting flux contours. In this figure, we clearly observe the minority electron current leakage from the QW into the p-side SCH region. It can also be seen that only a fraction of the minority electrons reaches the p-contact and the rest flow back to the QW and are reclaimed there.¹ In the latter case, a portion of them are attracted to the central region due to the spatial hole burning effect while the remaining flow to the outer region after encountering the oxide layers. Further simulations show that the density of the minority current is significantly reduced by grading the SCHs for the 2λ cavity VCSEL. This complicated behavior of the minority carriers cannot be modeled by simple rate equations. Under direct modulation, the accumulation and depletion of these carriers in the SCHs can induce a diffusion capacitance and this effective shunt capacitance can significantly dampen the dynamic response of VCSELs.

In Minilase, small signal modulation responses are obtained from time-dependent simulations. Two sets of such simulated modulation responses for 2λ cavity VCSELs with un-graded and graded SCHs, respectively, are plotted in Fig. 3. At each of the 3 simulated bias points, the response curve for the graded structure (solid line) is significantly improved in both the modulation bandwidth and the relaxation strength, compared with that for the un-graded structure (dashed

line). For example, an increase of ~ 3 GHz in the -3dB bandwidth can be seen for the structure with graded SCHs at $6 \times I_{th}$. We have also simulated the response curves for the un-graded, 2λ cavity device with the minority current artificially suppressed. In such cases, an artificial Richardson constant $\tilde{A}^* = \alpha A^*$, instead of its realistic value A^* , is used in Eq. (1) to regulate the amount of carrier leakage from the QW into the SCHs. For $\alpha < 1$, the minority current escaping from the QW is greatly reduced. Modulation responses for

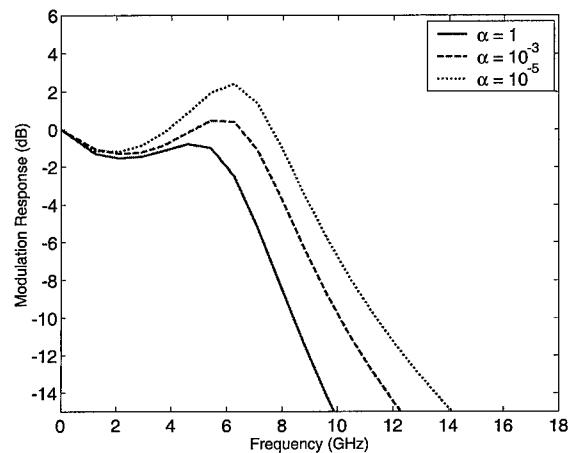


Figure 4. Simulated modulation responses for a un-graded, 2λ cavity VCSEL structure operating at $3 \times I_{th}$ with artificially suppressed minority current. In these simulations, the Richardson constant A^* is multiplied by a pre-factor, α , in the computation for the thermionic emission of the minority carriers.

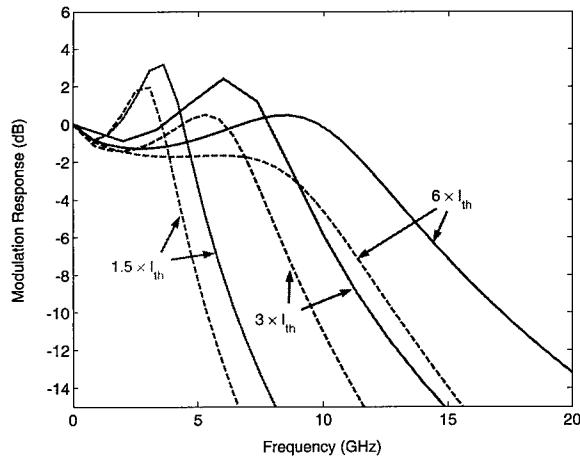


Figure 3. Simulated modulation responses for 2λ cavity VCSEL structures at current biases $1.5 \times I_{th}$, $3 \times I_{th}$ and $6 \times I_{th}$, respectively. Solid lines: 2λ cavity VCSEL with linearly graded SCHs. Dashed lines: 2λ cavity VCSEL with un-graded SCHs.

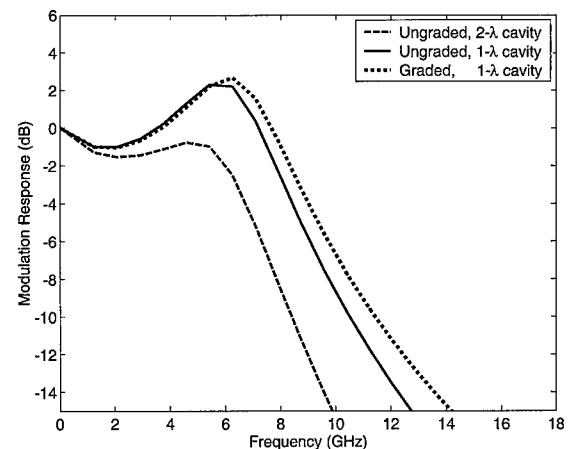


Figure 5. Simulated modulation responses at $3 \times I_{th}$ for 2λ cavity with un-graded SCHs, 1λ cavity with un-graded SCHs and 1λ cavity with graded SCHs, respectively.

$\alpha = 1, 10^{-3}, 10^{-5}$ at the current bias $3 \times I_{th}$ are plotted in Fig. 4, respectively. As α decreases, improvements in the modulation responses similar to those achieved by grading the SCHs can be seen in Fig. 4. This strongly suggests that the diffusion capacitance induced by the minority carriers is responsible for over-damping of the modulation responses of the un-graded, 2λ cavity VCSEL. However, for the practically always used 1λ cavity structure, the modulation response shows much reduced damping effects even with the un-graded SCHs, which is plotted as the solid line in Fig. 5. By comparing it to the response curve of the graded structure (dotted line in Fig. 5), we can also see that grading the SCHs of the 1λ cavity does not yield further improvements on the dynamic response. This result is quite different from what has been shown for the 2λ cavity in Fig. 3, where significant improvements to the responses are observed by grading the SCHs. It is hence indicated from these observations that the diffusion capacitance is much suppressed by shorter SCHs.

3. Conclusion

In conclusion, our Minilase simulations with the double oxide-confined VCSELs have revealed the complicated dynamic behaviors of minority carriers. Our simulations of the modulation responses indicate that a diffusion capacitance is responsible for a reduction of the modulation bandwidth. It is also shown that this diffusion capacitance is greatly dependent on the structure of the VCSELs and can be suppressed by careful device design. Our simulations demonstrate that, by grading the SCHs and reducing the thickness of the SCHs, the over-damping of the dynamic responses associated with the diffusion capacitance can be greatly reduced.

Acknowledgments

This project is supported by the Office of Naval Research and the National Science Foundation through the DesCartES center.

Note

1. In the simulations presented in this paper, we assume that the spontaneous emission is the dominant recombination mechanism for the carriers in the SCHs and the corresponding carrier lifetime is \sim nsec. Under this assumption, the fraction of the minority carriers that recombine in the SCHs is not significant. In a real device, however, the dark recombination of the minority carriers, particularly near the oxide interfaces, can be large.

References

- Grupen M. and Hess K. 1998. Simulation of carrier transport and nonlinearities in quantum well laser diodes. *IEEE Journal of Quantum Electronics* 34: 120–140.
- Hess K. 1988. *Advanced Theory of Semiconductor Devices*. Prentice Hall Press, Englewood Cliffs, New Jersey.
- Klein B., Register L.F., Hess K., Deppe D.G., and Deng Q. 1998. Self-consistent Green's function approach to the analysis of dielectrically apertured vertical-cavity surface-emitting lasers. *Applied Physics Letters* 73: 3324–3326.
- Lear K.L., Mar A., Choquette K.D., Kilcoyne S.P., Schneider R.P., and Geib K.M. 1996. High-frequency modulation of oxide-confined vertical cavity surface emitting lasers. *Electronics Letters* 32: 457–458.
- Nagarajan R., Ishikawa M., Fukushima T., Geels R.S., and Bowers J.E. 1992. High speed quantum-well lasers and carrier transport effects. *IEEE Journal of Quantum Electronics* 28: 1990–2007.
- Oyafuso F.A., Klein B.D., Register L.F., and Hess K. 2000. Fully coupled electrical and optical simulation of VCSELs. In: Choquette K.D. and Lei C. (Eds.), *Proceedings of SPIE Vol. 3946. Vertical-Cavity Surface-Emitting Lasers IV*, Bellingham, WA, pp. 108–116.
- Rideout W., Sharfin W.F., Koteles E.S., Vassell M.O., and Elman B. 1991. Well-barrier hole burnings in quantum well lasers. *IEEE Photonics Technology Letters* 3: 784–786.



Quantum Transport Simulation of Carrier Capture and Transport within Tunnel Injection Lasers

WANQIANG CHEN, XIN ZHENG AND LEONARD F. REGISTER

*Microelectronics Research Center, Department of Electrical and Computer Engineering,
The University of Texas at Austin, Austin, Texas 78758, USA*

MICHAEL STROSCIO

*Departments of Bioengineering, and Electrical and Computer Engineering, University of Illinois at Chicago,
851 S. Morgan St., Chicago, IL 60091, USA*

Abstract. Hot electron distributions within the active region of quantum well lasers lead to gain suppression, reduced quantum efficiency, and increased diffusion capacitance, greater low-frequency roll-off and high-frequency chirp. Recently, “tunnel injection lasers” have been developed to minimize electron heating within the active quantum well region by direct injection of cool electrons from the separate confinement region into the lasing subband(s) through a tunneling barrier. Tunnel injection lasers, however, also present a rich physics of transport and scattering, and a correspondingly rich set of challenges to simulation and device optimization. For example, a Golden-Rule-based analysis of the carrier injection into the active region of the ideal tunnel injection laser would suggest approximately uniform injection of electrons among the nominally degenerate ground quantum well states from the separate confinement region states. However, such an analysis ignores (via a random-phase approximation among the final states) the basic real-space transport requirement that injected carriers still must pass through the wells sequentially, coherently or otherwise, with an associated attenuation of the injected current into each subsequent well due to electron-hole recombination in the prior well. Transport among the wells then can be either thermionic, or, of theoretically increasing importance for low temperature carriers, via tunneling. Coherent resonant tunneling between wells, however, is sensitive to the potential drops between wells that split the energies of the lasing subbands and (further) localizes the electron states to individual wells. In this work such transport issues are elucidated using Schrödinger Equation Monte Carlo (SEMC) based quantum transport simulation.

Keywords: quantum transport, tunnel injection lasers, phonon scattering

1. Introduction

In conventional quantum well lasers, in principle, the time required for captured electrons to reach low energy states of the lasing subband via dissipative phonon interactions can be comparable to or larger than the lifetime of the carriers within the lasing subband prior to recombination, resulting in a hot electron distribution within the quantum wells (Bhattacharya 1998). Hot electron distribution, in turn, lead to gain suppression, reduced quantum efficiency, and increased diffusion

capacitance, greater low-frequency roll-off and high-frequency chirp (Bhattacharya 1998, Grupen and Hess 1997, 1998). Recently, “tunnel injection lasers” have been developed to minimize electron heating within the active quantum well region by direct injection of cool electrons from the separate confinement region into the lasing subband(s) through a tunneling barrier.

Tunnel injection lasers, however, also present a rich physics of transport and scattering, and a correspondingly rich set of challenges to simulation and device optimization. The quantum well region is

designed such that the lasing subbands of the quantum wells are nominally degenerate and, thus, delocalized among the wells. A Golden-Rule-based analysis of the carrier injection into the active region, which by design is phonon assisted, would suggest an approximately uniform injection of electrons into any number of quantum wells. However, such an analysis overlooks (in part, via a random-phase approximation among the final states) the basic real-space transport requirement that injected carriers still must pass through the wells sequentially, coherently or otherwise, with an associated attenuation of the injected current into each subsequent well due to electron-hole recombination in the prior well. The required transport among the wells then can be either thermionic, or, of theoretically increasing importance for low temperature carriers, phonon-assisted or coherent tunneling. Coherent resonant tunneling between wells, however, is sensitive to even small potential drops that split the energies of the lasing subbands and localizes the electron states to individual wells.

In this work these issues are addressed using Schrödinger Equation Monte Carlo (SEMC) based quantum transport simulation (Register 1998). SEMC provides a qualitatively and quantitatively accurate, non-perturbative, current conserving treatment of coherent transport and incoherent/phonon-mediated transport due to real scattering processes, including the dominant process of long-range polar-optical phonon scattering, and already has been used to study the effects of phase-coherence and phase-breaking on carrier capture by quantum wells (Register and Hess 1997).

2. Schrödinger Equation Monte Carlo

The SEMC method is described in detail in Register (1998); a brief summary is provided here. Phase breaking and energy dissipation within this Schrödinger Equation-based method are modeled via the exchange of probability among oscillator degrees of freedom within a many-body electron-phonon system just as in the true carrier-phonon scattering. For phonon scattering, a set of Schrödinger Equations is defined for the charge carrier corresponding to an "initial" state and many (e.g., 100 s or 1000 s of) "final" states separated from the initial state by the emission or absorption of one phonon. Coupling potentials between the initial and final states are provided by Monte Carlo sampling of the (spatial correlation functions of the) true carrier-phonon interactions. A probability source

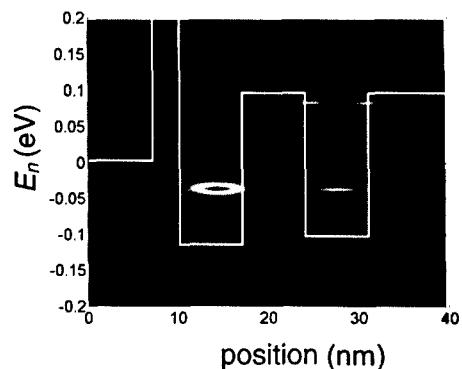


Figure 1. Carrier distribution (probability density) after the first scattering event as a function of position and well-normal component of energy for, in this and all following figures, a thermal distribution of electrons incident from the left, with the conduction band edge shown for reference.

to the initial state is provided by an open boundary in the carrier coordinates or coupling to a prior phonon state; probability sinks are provided by open boundaries in the carrier coordinates of both the initial and final states and/or, as required for bound final states in this work, complex "self-energy" potentials in the final states. This system, as depicted in Fig. 1(a) is solved self-consistently to find the many-body carrier-phonon wave-function from which any physical observable (transmission, reflection and capture probabilities, self-energies/scattering rates, currents in real-space or "phonon-space," etc.) can be obtained. Probability, energy and even phase information are inherently conserved with respect to full-many body system, but with respect to the carrier alone, the interaction is inelastic and phase breaking.

This procedure precisely emulates scattering, both real and virtual, in the true carrier-phonon system to first-order, and to higher orders within the accuracy of the estimated final-state self-energies. Scattering is neither local in position nor time. Indeed, the calculations of this work are time-independent (propagating) energy eigenstate calculations in the coupled carrier-phonon system. "Initial" and "final" only indicates the direction of probability current flow.

This basic procedure can be repeated sequentially to trace carriers through an unlimited number of scattering "events." The old initial state becomes the source, a new intermediate state is selected by Monte Carlo sampling from among the old final states according to the probability flow to/through the final states, and a new set of final states is generated each with its own complex self-energy potentials.

Table 1. Well/barrier width w , conduction band edge E_c , effective mass in units of m_e , and nonparabolicity γ for model tunnel injection laser systems with (A) high and (B) low interwell potential barriers.

Sep. conf.	Injection barrier	Leading well	Interwell barrier	Trailing well	Sep. conf.
(A) High interwell potential barriers					
w	NA	3	7	7	NA
E_c	0	1000	-117	100	-100
m^*	0.67	0.14	.047	.071	.047
γ	0.61	0.25	1.02	0.57	1.02
(B) Low interwell potential barriers					
w	NA	3	7	7	NA
E_c	0	1000	-101	0	-87
m^*	.067	0.14	.047	.067	.047
γ	0.61	0.25	1.02	0.61	1.02

3. Model Tunnel Injection Laser Structures

The design objective of the tunnel injection laser is to deliver, in particular, “cool” electrons uniformly among multiple quantum wells for recombination with holes. The goal of this work is to study the essential physics of this process within a quantum transport formalism, and compare the results with the expectations from simpler approaches. To this end, the quantum well/barrier structure to be simulated has been simplified by reducing the number of wells to one leading well and one following well, between which disparities in the carrier densities and capture rates and the reasons for those disparities may be more readily identified. Two structures are considered, modeled after those discussed in Bhattacharya (1998), as defined in Tables 1(A) and (B), respectively. Both of these structures are designed to have nominally degenerate ground state energies in the two wells 39 meV below the conduction band edge of the electron injection side of the separate confinement region. The only significant difference between the two structures is the height of the barrier between the two wells; one equal to the height of the band edge on the hole injection side of the separate confinement region, the other equal to the height of the band edge on the electron injection side of the separate confinement region. Polar optical phonon scattering due to GaAs bulk modes was considered within the active region in this work. Consideration of the full spectrum of interface and confined phonon modes (Yu *et al.* 1997) is planned and ultimately necessary.

4. Simulation Results and Discussion

As noted above, a Golden Rule-based analysis would suggest uniform injection of electrons into the ground

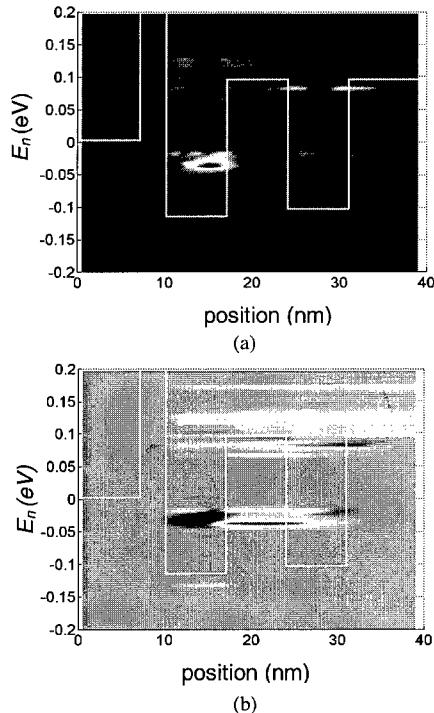


Figure 2. (a) The current flow, which is localized primarily to the leading well, from the initial/incident electron state to the final/captured electrons states (in “phonon-space”) as a function of the position and well-normal component of energy of the captured electrons. (b) The subsequent resonant real-space tunneling current from the first well to the second.

states of the two wells from low-energy incident electrons. However, as shown in Fig. 1, there is a decided segregation of the charge after the initial capture/scattering event toward the leading well, and a significant fraction of the charge captured in the second well enters hot from higher energy incident electrons into the excited state of that well. Those electrons that are found in the ground state of the second well at low energies after one scattering event get there by tunneling from the first well *after* the scattering event, as shown in Fig. 2. Over time, the interwell tunneling will reduce but not eliminate this segregation, as shown in Fig. 3(a). However, if a small additional voltage bias is applied, 26 mV across the active region for this work, the resonant tunneling process is greatly attenuated resulting in further segregation of the carriers toward the leading well, as shown in Fig. 3(b).

Reducing the barrier height for the second structure eliminates much of this problem, however, by allowing penetration of the incident carrier wave-functions all of the way across the active region, leading to a nearly

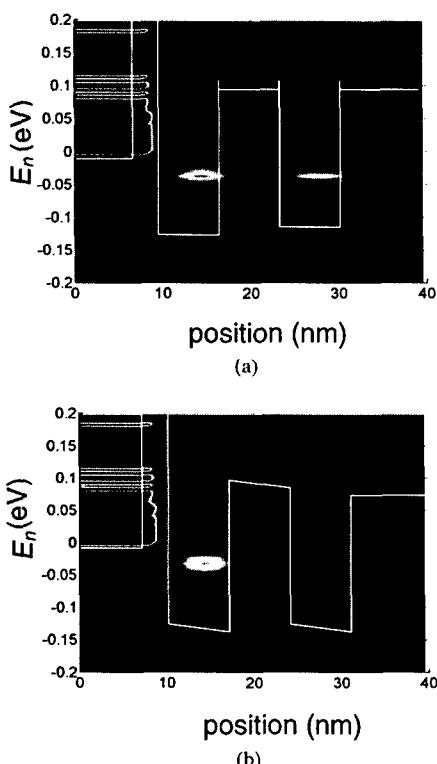


Figure 3. The cumulative probability densities within the wells through 10 scattering events as a function of position and well-normal component of energy (a) at flatband and (b) under additional bias.

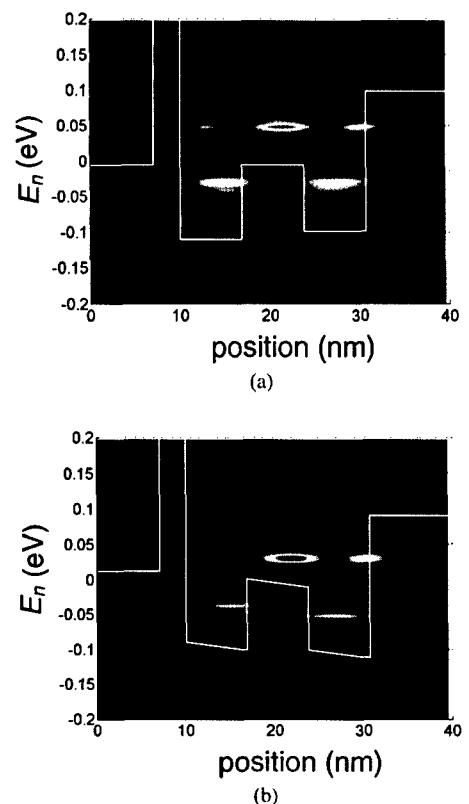


Figure 4. Carrier distribution (probability density) after the first scattering event as a function of position and well-normal component of energy (a) at flat-band, and (b) under additional bias (26 mV across active region).

uniform distribution of electrons between the two wells after the initial capture process (and subsequently) at flat band, as shown in Fig. 4(b), or under additional bias, as shown in Fig. 4(b). We note that in these simulations, incident, electrons enter the low barrier system with somewhat more energy than they enter the high barrier system, as seen in Fig. 5. However this result is a likely artifact of the scale of the system; for larger, more realistic numbers of wells, lower energy quasi-confined states should exist within the active region of the low barrier system to tunnel into.

In one regard, both systems performed well (at least for a two well system) but particularly the low barrier system. As also seen in Fig. 5, the fraction of carriers that leak beyond the well—resulting in diffusion capacitance and dark current—is quite low because of the offset between the conduction band-edge in the separate confinement region on the electron injection side and that on the hole injection side. Similarly, the tunnel barrier to electron injection should minimize

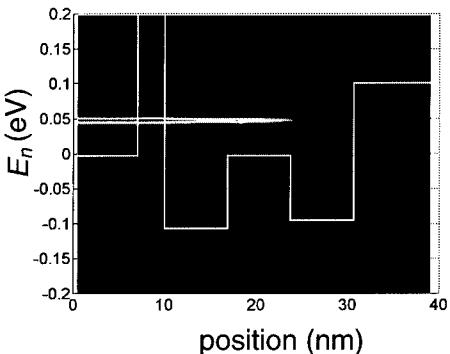


Figure 5. Real-space current flow into and, to a small extent, beyond the active region as a function position and well-normal component of energy.

hole transport beyond the active region that can be more significant than for electron transport as a result of thermionic emission between wells. Further, the tunneling barrier may serve to allow hot electrons injected from the cladding layer more time to cool within the separate confinement region before entering the active region. The cost of the barrier, however, is greatly reduced electron capture efficiency.

5. Conclusion

A preliminary study of transport in tunnel injection lasers has been performed. It has been demonstrated that tunnel injection lasers can offer advantages over more conventional lasers by, as intended, lowering the carrier injection energy and by, in addition, reducing leakage

currents. However, it has also been demonstrated that a Golden-Rule analysis of capture can be misleading, and that interwell transport may be quite sensitive to the voltage drops between wells. In order to make direct comparisons with experiments, comprehensive modeling should address these effects—although not necessarily requiring as rigorous a transport approach once the essential physics has been identified—and others. For example, although designed with electron transport in mind, these systems also offer advantages for hole transport as suggested above. In addition, the richer energy spectrum of the full set of phonon modes (Yu *et al.* 1997) as compared to that of a single mode used in this preliminary work, may effect and perhaps enhance carrier capture and interwell transport (Yu *et al.* 1997).

Acknowledgments

This work was supported by the U.S. Army Research Office and Battelle.

References

- Bhattacharya P. 1998. Int. J. High-Speed Electronics and Systems 9: 847 and references therein.
- Grupen M. and Hess K. 1997. Appl. Phys. Lett. 70: 808.
- Grupen M. and Hess K. 1998. J. Quantum Electronics 34: 120.
- Register L.F. 1998. Int. J. High-Speed Electronics and Systems 9: 251.
- Register L.F. and Hess K. 1997. Appl. Phys. Lett. 71: 1222.
- Yu SeGi *et al.* 1997. J. Appl. Phys. 82: 3363 and references therein.



Modeling of Semiconductor Optical Amplifiers

ANDREA REALE AND PAOLO LUGLI

INFM and Department of Electronic Engineering University of Roma “Tor Vergata”, I-00133 Roma, Italy

Abstract. We present a tight-binding analysis of the polarization dependence of GaAs δ -strained semiconductors optical amplifiers. We explain how thin strained GaAs layers embedded in a lattice-matched InGaAsP/InGaAs quantum well can be used to achieve polarization insensitive optical amplification. We describe also the interaction between pulse propagation and gain compression within a pump-probe excitation in polarization insensitive MQW-SOA. Another important non-linear effect studied is Four Wave Mixing (FWM) on the pulse propagation in the active region of SOAs. Our model successfully predicts operation of optical data sampling using FWM interaction between a signal bit stream and an optical clock.

Keywords: semiconductor optical amplifier, tight binding, polarization independence, cross gain modulation, four wave mixing, format conversion

1. Introduction

In this paper, first we study with the tight-binding method the optical amplification/absorption of a δ -strained semiconductor amplifier, which has been shown to be a very promising structure for optical communication systems operating at 1.55 μm (Seiferth *et al.* 1997).

We then present the model of gain compression mechanisms in a polarization insensitive MQW-SOA, and we compare our results to existing experimental data. The description of pulse propagation allow us to determine the light-carrier interaction consistently along the direction of the propagation axis. This is an important feature to model the non-linear operation of Traveling Wave Amplifiers (TWA).

In order to describe also the FWM non-linearity in SOAs, we discuss the theoretical model for the propagation of short optical pulses in the presence of FWM interaction. The simulation results for format conversion in a multiplexer scheme for OTDM systems are also shown.

2. The Tight-Binding Method Applied to Polarization Independent Semiconductor Optical Amplifiers

Optical properties can be easily calculated within the tight-binding scheme *without* introducing new fitting parameters (Di Carlo *et al.* 1996). If we consider a linear polarization of the light along the i -th axis, the absorption coefficient can be written as

$$\alpha(\omega) = \frac{4\pi^2}{nc\omega S} \sum_{E, E', \mathbf{k}_\parallel} [f(E) - f(E')] \cdot \delta(\hbar\omega + E - E') \cdot \left| \left\langle E, \mathbf{k}_\parallel \left| \sum_{\alpha, m} J_i(\alpha, m) \right| E', \mathbf{k}_\parallel \right\rangle \right|^2 \quad (1)$$

where n is the refractive index and c the speed of light. Here, S is the transverse area of the primitive cell, E and E' are the initial and final electron energies, respectively, $\hbar\omega$ is the photon energy and J_i the current operator (Graf and Vogl 1995). The basis set for the evaluation of the current operator is

given by the system wave functions $|E, \mathbf{k}_\parallel\rangle$. Strain is included in the tight-binding model by scaling the hopping matrix elements (Harrison 1980).

The numerical implementation of the TB approach is of crucial importance. By itself, the method is computationally quite heavy since the diagonalization of very large matrices is needed. In order to speed up the calculations, we have introduced a hybrid method to diagonalize the tight-binding Hamiltonian which uses a standard (LAPACK (Anderson 1992) routine to calculate eigenvalues and an inverse iteration scheme (Press *et al.* 1986) to calculate eigenvectors.

The reference structure for our study consists of 153 Å wide (52 monolayers) $\text{In}_{0.533}\text{Ga}_{0.467}\text{As}$ quantum well surrounded by $\text{In}_{0.74}\text{Ga}_{0.26}\text{As}_{0.56}\text{P}_{0.44}$ barriers, lattice matched to an InP substrate. We investigate the optical matrix elements when 3 monolayers (ML) of InGaAs in the middle of the well are replaced by tensile strained GaAs.

When δ -strain is present the first light-hole level lifts up in energy, as discussed in the previous section, while the first heavy-hole level shifts down, leading to a band

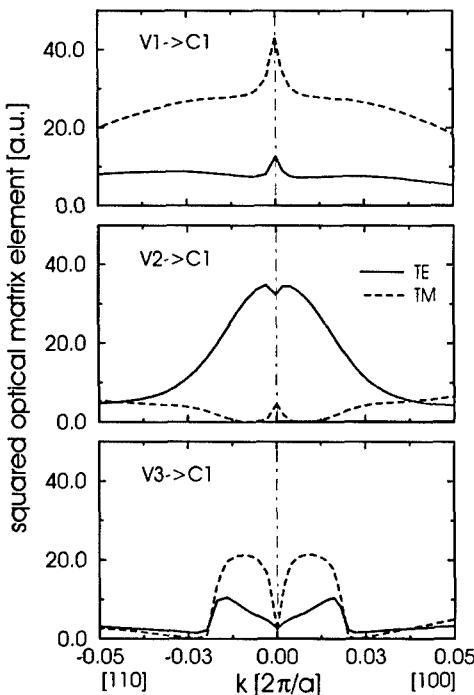


Figure 1. Squared optical matrix element as a function of the inplane k vector along [110] and [100] directions for a MQW-SOA with δ -strain. The contribution of each individual transition is distinguished.

degeneration at $\mathbf{k}_\parallel = 0$. The character of these states at zone center can be deduced by looking at the squared optical matrix elements (Fig. 1). We notice that the first valence band has a light hole character (first LH level), while the second valence band has a heavy-hole character (first HH level). Very interesting is the third valence band (second HH band), where the transition to C1 presents at $\mathbf{k}_\parallel = 0$ a TE contribution. This transition, which cannot be accounted in the $\mathbf{k} \cdot \mathbf{p}$ -EFA model, follows from band mixing at $\mathbf{k}_\parallel = 0$. The same mixing effect is responsible for the TM polarized $V2 \rightarrow C1$ transition at zone center.

3. Model of Carrier Dynamics during Pulse Propagation

In the description of carrier dynamics and propagation effects in MQW-SOA our approach explicitly includes the main transport mechanisms across the active region: exchange of carriers between the different QWs, exchange between QW and Separate Confinement Heterolayers (SCH) surrounding the QWs, and carrier injection from the SCH. We point out that the dynamics of carriers and the interaction with light are described in a coarse time-space grid.

The propagation of light pulses through an optically active medium as a function of time t and of the position z along the active region confinement guide is described by a set of coupled partial differential equations.

$$\frac{dS}{dz} = g(N) S - \frac{1}{v_g \tau_p} S + \frac{1}{v_g} \beta R_{sp} \quad (2)$$

$$\frac{\partial N_i}{\partial t} = \eta_{inj} \frac{I}{qL} - R(N_i) - R_{st}. \quad (3)$$

In Eq. (2) z is the propagation direction, $S = S(t, z)$ the total photon concentration, $N_i = N_i(t, z)$ the carrier density in the i -th well, v_g the group velocity, $g(N)$ the gain-carrier density relationship, τ_p photon lifetime, β the coupling factor of the spontaneous emission $R_{sp}(N_i)$ in the i -th well with the main propagating mode. In Eq. (3) η_{inj} is the injection efficiency, I the injected current, q the electron charge, L the thickness of the particular layer considered, $R(N_i)$ the recombination term accounting for trap-related, spontaneous and Auger recombinations (including also transport processes such tunneling or thermionic emission) and R_{st} accounts explicitly for stimulated recombinations.

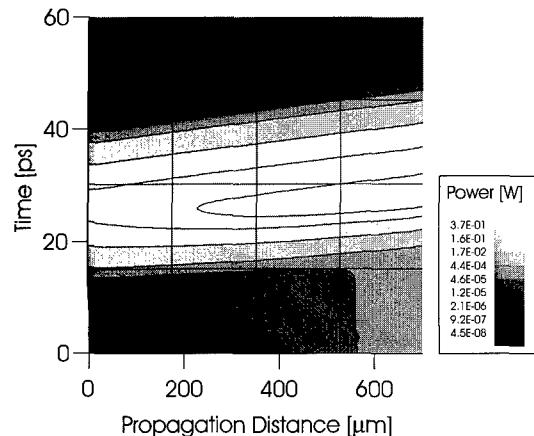


Figure 2. Pump amplification along the SOA waveguide for a current bias of 150 mA.

The gain versus carrier density relation is given by

$$g(N) = \begin{cases} G_0 \ln\left(\frac{N}{N_{th}}\right); & N \geq N_{th} \\ -\alpha; & N < N_{th}; \end{cases} \quad (4)$$

The parameters concerning the present propagation scheme are indicated elsewhere (Reale, Di Carlo and Lugli 2001). Figure 2 describes the propagation of a gaussian pulse (FWHM = 8 ps) injected into the device. The grey scale intensity reproduces the light intensity inside the semiconductor slab, the lighter grey scale corresponding to the higher intensities. The time and spatial coordinates are given respectively by the vertical and horizontal axis.

The modulation of the carrier density is described in Fig. 3. The time scale now is much longer than the one in Fig. 2, as the density modifications induced by the pump pulses (whose duration is less than 10 ps) are recognizable only at nanosecond delay times. The reason of such different behavior is that carrier depletion takes place (almost instantaneously) by stimulated recombination, while the recovery of stationary conditions is limited by the slower non radiative recombination mechanisms.

The process of carrier recombination described by Fig. 3 results in a negative spike in the probe output as observed in the experiment of Reale *et al.* (1999) and as illustrated in the simulation result of Fig. 4. The shape of this “dark pulse” at the output is asymmetric, due to the fact that the falling portion slope reproduces the fast stimulated recombination process induced by the pump pulse, while the rising part is limited by the non

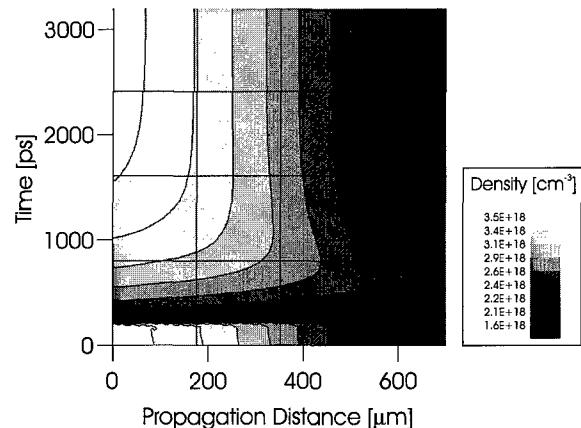


Figure 3. Carrier density along the SOA waveguide for a current bias of 150 mA.

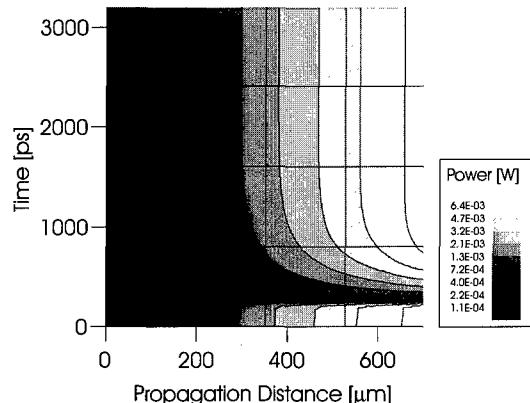


Figure 4. Probe amplification along the SOA waveguide for a current bias of 150 mA.

radiative recombination time controlling the recovery to the stationary level of the carrier population.

4. Theoretical Model of the Four Wave Mixing Interaction

Modeling the FWM process requires the knowledge of the microscopic phenomena leading to the system non linear response (Yariv 1997). In SOAs we are concerned primarily with gain dynamics resulting from interband and intraband processes (Guekos 1998).

Equations (2) and (3) describe the equilibrium carrier dynamics in the presence of a propagating electromagnetic field. In order to include CH and SHB effects, one has to model the variation $\Delta n_{shb} = n - n_{Fermi}$, $\Delta n_{ch} = n_{Fermi} - n_{Fermi,eq}$ in the local carrier density due to these

non equilibrium processes. Here n is the local carrier density, n_{Fermi} is the local carrier density evaluated with the heated Fermi distribution, $n_{Fermi,eq}$ is the local carrier density in equilibrium with the lattice evaluated with the thermalized Fermi distribution.

If the pulses propagating in the amplifier are longer than the characteristic times of SHB and CH, one can assume that Δn_{shb} and Δn_{ch} are in quasi-equilibrium with the instantaneous field value. Under this assumption one has

$$\Delta n_{shb} = -\frac{v_g}{a_N} \epsilon_{shb} g S \quad (5)$$

$$\Delta n_{ch} = -\frac{v_g}{a_N} \epsilon_{ch} g S. \quad (6)$$

where ϵ_{shb} is the gain compression factor due to SHB, ϵ_{ch} the gain compression factor due to CH, and a_N/v_g is the gain cross section.

The explicit contribution of the refractive index modulation on the phase changes along the waveguide can be described in terms of variations of the field wavevector $k = 2\pi n/\lambda$ as follows (Guekos 1998)

$$\frac{\partial \Phi}{\partial z} = -\frac{1}{2} \Gamma \alpha_N \frac{\partial g}{\partial N} \Delta N - \frac{1}{2} \Gamma \alpha_T \frac{\partial g}{\partial N} \Delta n_{ch} \quad (7)$$

where ΔN represents the variation of the carrier density from the stationary value in absence of optical input and the so called *alpha factors* (Guekos 1998).

A proper filtering of the spectrum of the total electric field $\mathbf{E}(t, z)$ calculated at the output facet ($z = L$) permits the extraction of the generated conjugated signal $\mathbf{E}_c(t)$. To perform this operation, we calculate the Fourier transform of $\mathbf{E}(t, L)$ and model the Optical Band-Pass Filter (OBPF) with a raised cosine function of spectral width $\Delta\omega_{OBPF}$. The filter output can then be transformed backward in order to obtain the time domain representation of the conjugated signal.

One of the potential applications of FWM in ultrafast optical network is related to the fact that FWM acts as a logical AND operation between pump and signal fields (Kamatani and Kawanishi 1996, Nasset *et al.* 1994). Thus, it can then be used as a pulse reshaping and format converter tool for optical systems operating up to tens of Gbit/s. In particular, it might be possible to operate a format conversion between long NRZ pulses to short duration RZ pulses. We considered 400 ps long NRZ gaussian pulses at 2.5 Gbit/s, having an average power for the on state of -10 dBm and an extinction ratio between the on and the off state of 10 dB. The local clock RZ source is a gaussian pulse with a duration of 25 ps and a pulse energy of 10 fJ. The RZ pulse train is synchronized with the maximum of the NRZ pulses to maximize the FWM interaction.

Figure 5 depicts the intensity of the total electric field at the input of the SOA (lower trace) and at the output (upper trace) as a function of time. The NRZ bit sequence is 10010 and the beating pattern observed

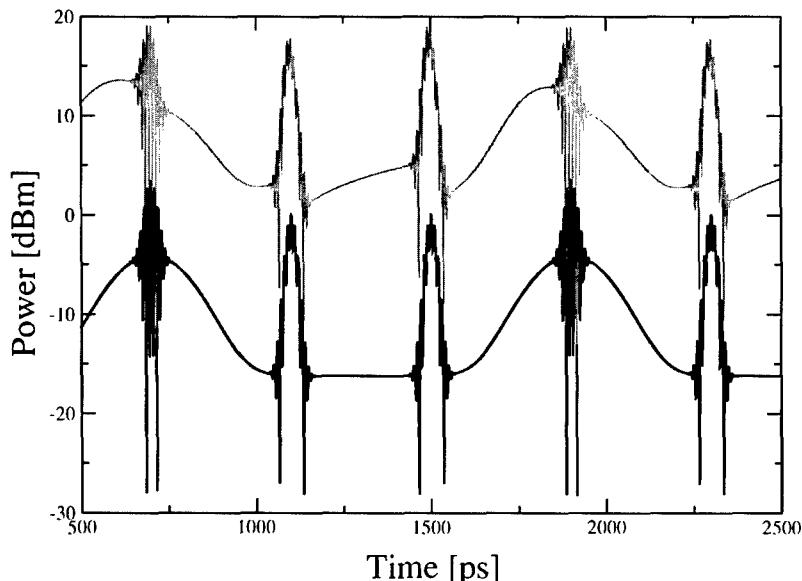


Figure 5. Total input and output intensity in FWM interaction between NRZ and RZ pulse train.

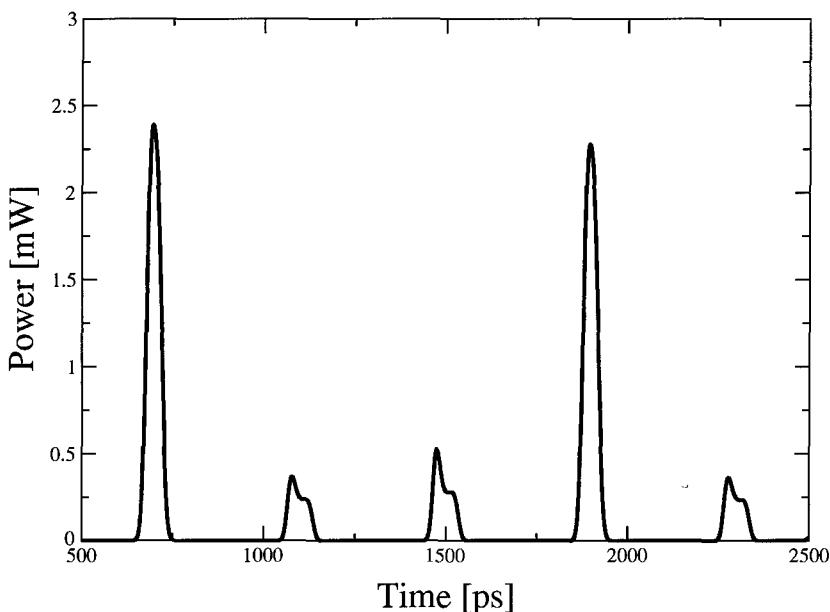


Figure 6. Conjugate signal showing the sampled bit sequence of the input NRZ pulses.

when both NRZ and RZ signal are present is due to the presence of the beating factor $e^{j\Delta\Omega t}$ in the total photon density $S(t, z)$. From input to output it can also be noted how the NRZ pulses get distorted due to gain saturation in the SOA.

It can be observed that the conjugated spectra can be filtered out once a proper filtering is performed around the central angular frequency $\omega_c = \omega_p - \Delta\Omega$. The backward transformation of the conjugate spectra shown in Fig. 6 clearly indicates the logical AND operation associated to FWM. The clock pulses are sampling the logical level of the NRZ pulses. The extinction ratio of the conjugated signal is less than 10 dB. However the present work demonstrates the possibility to investigate the conditions for which such extinction ratio can be raised to higher values. This is of fundamental importance to investigate the practical feasibility of pulse reshaping—format conversion for multiplexing in OTDM systems (Gosset and Hua Duan 2001, Jiang *et al.* 1999).

5. Conclusions

The influence of a delta-strain on the modal absorption/gain characteristic of a semiconductor optical amplifier have been studied by means of a tight-binding calculation. We have presented a model for studying gain

compression and propagation dynamics in Traveling Wave Multiple Quantum Well SOAs. We have presented a theoretical study of FWM phenomena in Semiconductor Optical Amplifiers. Our results demonstrate that FWM can be usefully used as a pulse reshaping and format converter tool. This can be of fundamental importance when raising the speed of optical systems up to tens of Gbit/s.

Acknowledgments

We acknowledge the support of Progetto Finalizzato CNR-Madess II and of MURST under the project “Sistemi e Tecniche per OTDM ad Altissimo Bit Rate”.

References

- Anderson E. 1992. LAPACK User's Guide, SIAM, Philadelphia.
- Di Carlo A., Pescetelli S., Paciotti M., Lugli P., and Graf M. 1996. Solid State Comm. 98: 803.
- Gosset C. and Hua Duan G. 2001. IEEE Phot. Tech. Lett. 13: 139.
- Graf M. and Vogl P. 1995. Phys. Rev. B 51: 4940.
- Guekos G. (Eds.). 1998. Photonic Devices for Telecommunications. Springer-Verlag, Berlin.
- Harrison W.A. 1980. Electronic Structure and the Properties of Solids. Freeman, San Francisco.
- Jiang L., Ippen E.P., Diez S., Hilliger E., Schmidt C., and Weber H.G. 1999. Summaries of Papers Presented at the CLEO'99, LEOS, Piscataway, p. 444.

- Kamatani O. and Kawanishi S. 1996. *J. Ligth. Tech.* 14: 1757.
Nesset D., Tatham M.C., Westbrook L.D., and Cotter D. 1994. *El. Lett.* 30: 1938.
Press W.H., Flannery B.P., Teukolsky S.A., and Vetterling W.T. 1986. *Numerical Recipes*. Cambridge University Press.
Reale A., Di Carlo A., and Lugli P. 2001. *IEEE J.Sel.Top.Q.El* 7: 293.
Reale A., Di Carlo A., Lugli P., Campi D., Cacciatore C., Stano A., and Fornuto G. 1999. *IEEE J.Q.El.* 35: 1697.
Seifert F., Johnson F.G., Merritt S.A., Fox S., Whaley R.D., Chen Y.J., Degenais M., and Stone D.R. 1997. *IEEE Photon. Technol. Lett.* 9: 1340.
Yariv A. 1997. *Optical Electronics in Modern Communications*. Oxford University Press, New York.



Hybrid LSDA/Diffusion Quantum Monte-Carlo Method for Spin Sequences in Vertical Quantum Dots

P. MATAGNE

*Beckman Institute for Advanced Science & Technology, University of Illinois at Urbana-Champaign,
405 N. Mathews Avenue, Urbana, Illinois 61801, USA*

T. WILKENS

Department of Physics, University of Illinois, Urbana, Illinois 61801, USA

J.P. LEBURTON

*Beckman Institute for Advanced Science & Technology, University of Illinois at Urbana-Champaign,
405 N. Mathews Avenue, Urbana, Illinois 61801, USA*

R. MARTIN

Department of Physics, University of Illinois, Urbana, Illinois 61801, USA

Abstract. We present a new hybrid Diffusion Quantum Monte-Carlo (DQMC)/Local Spin Density Approximation (LSDA) method, to compute the electronic structure of vertical quantum dots (VQD). The exact many-body electronic configuration is computed with a realistic confining potential. Our model confirms the atomic-like model of 2D shell structures obeying Hund's rule already predicted by LSDA.

Keywords: quantum dots, Diffusion Quantum Monte-Carlo, LSDA, Hund's rule

1. Introduction

Spin effects, and their possible manipulation by electric gating in quantum dots have received significant attention because of the new physics associated with few spin systems and their potential applications in quantum information processing. Various models have been used to approximate the many-body Schrödinger equation: exact diagonalization (ED) (Ezaki, Mori and Hamaguchi 1997, Imamura *et al.* 1995), quantum Monte-Carlo (QMC) (Bolton 1994, Pederiva, Umrigar and Lipparini 2000), density functional theory (DFT) (Macuccci, Hess and Iafrate 1993, Stopa 1996, Lee *et al.* 1998, Matagne *et al.* 2000) and Hartree-Fock (HF) (Yannouleas and Landman 1999). These different models, however, have predicted contradictory phenomena ranging from Wigner localization (Yannouleas

and Landman 1999), spin density waves (Yannouleas and Landman 1999) and atomic-like properties such as shell filling with Hund's rule (Ezaki, Mori and Hamaguchi 1997, Stopa 1996, Lee *et al.* 1998, Matagne *et al.* 2000).

In this paper we present a hybrid Local Density Approximation (LSDA)/Diffusion Quantum Monte Carlo (DQMC) method for simulating the electronic configurations of realistic vertical quantum dots (Tarucha *et al.* 1996). This approach has the unique advantage of combining the flexibility of the LSDA for modeling the device features of the quantum dot with the accuracy of the DQMC for computing the ground state of the many-body system. The method computes the three-dimensional (3D) self-consistent confining potential from Poisson and Kohn-Sham equations with the realistic device structure comprising hetero-barriers and

doping regions with boundary conditions on the electric potential deduced from the external bias applied to the gate. The exact many-body electronic configurations at each gate voltage are then simulated by using DQMC with the realistic 3D potential. We show that the relative error between the total energy computed by LSDA and QDQC never exceeds 7%. Moreover, LSDA and DQMC are in excellent agreement for the spin configuration that leads to the lowest energy, for any number, N , of electrons between 2 and 16, and confirm the atomic-like model of 2D shell structures, obeying Hund's rule for open shells.

2. Structure Description and Device Operation

Figure 1(a) shows a schematic diagram of a CVQD similar to the device investigated by Tarucha *et al.* (1996). The quantum dots reside in devices fabricated from a double barrier heterostructure (DBH) consisting of an undoped 12 nm $\text{In}_{0.05}\text{Ga}_{0.95}\text{As}$ well and undoped 9 nm and 7.5 nm $\text{Al}_{0.22}\text{Ga}_{0.78}\text{As}$ barriers (Fig. 1(a)). The source and drain leads on both sides of the DBH are made of $n^+\text{GaAs}$. The diameter, $d = 2R$ of the measured mesas is 0.5 μm , but the effective dot radius is 50 nm.

3. Approximations for the Many-Body Problem

In order to obtain the electronic and spin properties of the structure described above, we are concerned with solving the non-relativistic, time independent, many-body Schrödinger equation under the Born-Oppenheimer and effective mass approximations,

$$\left[-\frac{\hbar^2}{2m^*} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N V_{ext}(\mathbf{r}_i) + \sum_{i,j=1; i \neq j}^N \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \times \Psi(\mathbf{R}) = E\Psi(\mathbf{R}), \quad (1)$$

where N is the number of electrons, m^* is the electron effective mass, $\mathbf{R} \equiv (\mathbf{r}_1, \dots, \mathbf{r}_N)$ and $V(\mathbf{r}_i)$ is the external potential energy operator for the i th electron. Approximations for simplifying Eq. (1) occur at three levels: (i) the confining potential $V_{ext}(\mathbf{r})$; (ii) the problem dimensionality; (iii) the many-body electron-electron interaction.

3.1. The Electron Confining Potential and Dimensionality

Many authors (Ezaki, Mori and Hamaguchi 1997, Bolton 1994, Pederiva, Umrigh and Lipparini 2000,

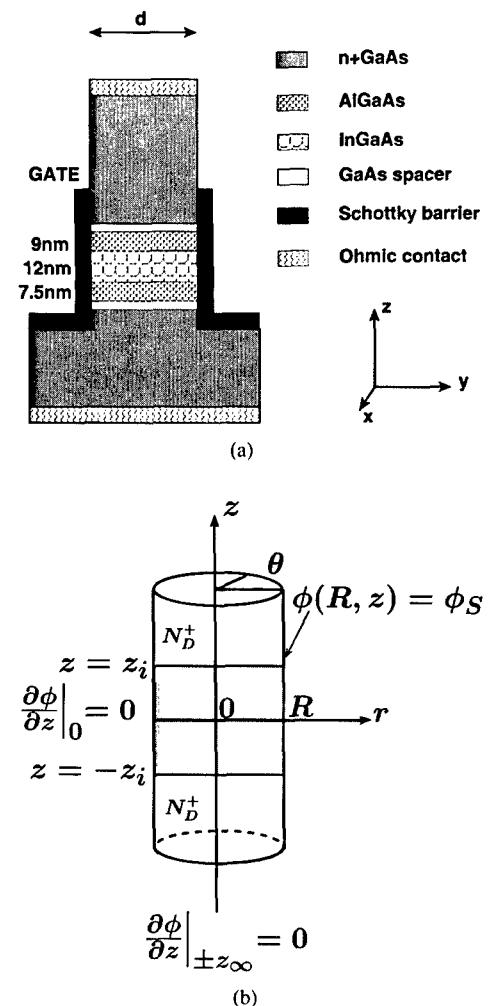


Figure 1. (a) Schematic diagram of a cylindrical vertical quantum dot tunneling heterostructure showing the different semiconductor layers. (b) Cylindrical charge model for the CVQD structure with boundary conditions.

Lee *et al.* 1998, Yannouleas and Landman 1999) use the parabolic approximation for V_{ext} in the 2DEG ($x-y$) plane: $V_{ext}(r) = (1/2)\omega^2 r^2$. However, by solving Poisson equation on the cylindrical domain shown on Fig. 1(b)), we have shown (Matagne and Leburton submitted) that the confining potential in the $x-y$ plane V_{ext} is not purely parabolic and that higher order terms lift all the accidental degeneracies that would appear in the eigenlevel spectrum with the parabolic approximation. In addition, we have also shown that, as V_G is swept, i.e., as the number of electrons in the dot increases, the quadratic term in the confining potential which determines the oscillator frequency decreases. Moreover, Rontani *et al.* (1999) have shown that, by solving

the problem in the $x-y$ plane only, and neglecting the vertical direction, the carrier localization is overestimated and often lead to an inadequate description of the Coulomb interaction between electrons. Thus, in order to compute the confining potential accurately, a 3D Poisson equation has to be solved with realistic device structure and appropriate boundary conditions deduced from the applied gate bias.

3.2. The Electron-Electron Interaction

Many-body methods, such as ED and QMC, compute the electron-electron interaction exactly. In particular, the many-body wave functions explicitly incorporate electronic correlation. Unfortunately, the phase space grows exponentially with the number of electrons, which restricts the non-stochastic many-body technique (ED) to a small number of electrons ($N = 12$ (Imamura *et al.* 1995)). QMC, however, scales the exponential complexity down to N^3 while DFT methods, with their shortcoming, are computationally less expensive and therefore more suitable to be coupled self-consistently to Poisson equation. To the best of the author's knowledge, the many-body methods are always used with an ideal parabolic confining potential. In the next section, we present a method that combines the advantages of LSDA and DQMC.

4. The Hybrid LSDA/DQMC Method (LSDA/DQMC)

The flowchart of the LSDA/DQMC Method can be stated as follow:

(a) Solve Kohn-Sham Equations,

$$\left[-\frac{\hbar^2}{2} \nabla \left(\frac{1}{m^*} \nabla \right) - q(\phi_{ext} + \phi_{ion} + \phi_H + \phi_{off}) + \mu_{xc}^{\uparrow(\downarrow)}[n] \right] |\phi_i^{\uparrow(\downarrow)}\rangle = E_i |\phi_i^{\uparrow(\downarrow)}\rangle$$

where $\phi(\mathbf{r}) = \phi_{ext} + \phi_{ion} + \phi_H$ is the electrostatic potential which consists of three contributions: ϕ_{ext} is the potential due to external applied bias, ϕ_{ion} is the potential resulting from ionized donors and ϕ_H is the Hartree potential accounting for repulsive electron-electron interactions. ϕ_{off} is the conduction band offset between different materials (Matagne and Leburton submitted).

(b) Compute the electron density

$$n(\mathbf{r}) = n^\uparrow(\mathbf{r}) + n^\downarrow(\mathbf{r}) = \sum_{i=1}^{N_\uparrow} |\psi_i^\uparrow(\mathbf{r})|^2 + \sum_{i=1}^{N_\downarrow} |\psi_i^\downarrow(\mathbf{r})|^2$$

where $N_\uparrow(N_\downarrow)$ is the number of spin up (down) electrons.

(c) Solve Poisson Equation

$$\nabla(\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -\rho(\mathbf{r})$$

where $\epsilon(\mathbf{r})$ is the position dependent permittivity and $\rho(\mathbf{r})$ is the total charge density which is given by $\rho(\mathbf{r}) = q(N_D^+(\mathbf{r}) - n(\mathbf{r}))$, where $n(\mathbf{r})$, $N_D^+(\mathbf{r})$ are the electron and ionized donors densities respectively, at the position \mathbf{r} .

(d) Compute the exchange correlation potential (Wang and Chou 1993)

$$\mu_{xc}^{\uparrow(\downarrow)} = \frac{d(n\varepsilon_{xc}[n])}{dn^{\uparrow(\downarrow)}}$$

(e) Go back to (a) until convergence is achieved.

Equations (a) and (c) are discretized by the finite element method of which the detailed formulation has been published elsewhere (Matagne *et al.* 2000).

Now, a first guess Ψ_0 for the many-body wave function Ψ is constructed as a linear combination of Slater determinants of the single-particle wave functions ψ_i and the external potential energy part of the many-body hamiltonian takes the realistic form $V_{ext} = -q(\phi_{ion} + \phi_{ext} + \phi_{off})$ which is extracted from ϕ by Green's function. The many-body problem can then be solved by DQMC method (Wilkens 2001).

5. Results

Figure 2 shows the various energy contributions to the electronic system in the dot computed with LSDA and LSDA/DQMC. $\langle KE \rangle$, $\langle PE_{int} \rangle$ and $\langle PE_{ext} \rangle$ are the quantum mechanical average of the kinetic energy, internal potential energy and external potential energy, respectively. The internal potential energy includes the energy contributions due to electron-electron interaction while the external potential energy include the energy contribution due to the interaction between the

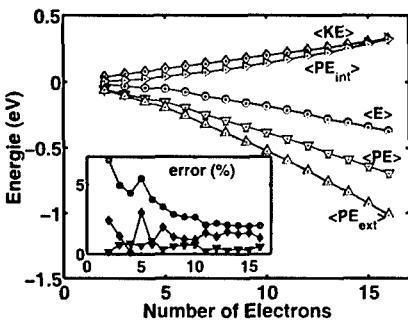


Figure 2. Comparison between the total energy $\langle E \rangle$, the kinetic energy $\langle KE \rangle$, the total potential energy $\langle PE \rangle$, the internal potential energy $\langle PE_{int} \rangle$ (Hartree + exchange) and the external potential energy $\langle PE_{ext} \rangle$ between LSDA (black) and DQMC (white) as a function of the number of electrons. Inset: relative error $|(DQMC - LSDA)|/DQMC|$ between LSDA and DQMC.

electrons and the external potential V_{ext} . Of course, $\langle PE \rangle = \langle PE_{int} \rangle + \langle PE_{ext} \rangle$ denotes the total potential energy and $\langle E \rangle = \langle KE \rangle + \langle PE \rangle$ is the total energy. The agreement is remarkable since the relative LSDA error with respect to LSDA/DQMC (inset) never exceeds 7% and converges to 2% when the number of electrons becomes large.

Moreover, we have computed all the electronic configurations for $N = 2$ to 16. In each case, LSDA and LSDA/DQMC agree on the ground state configuration. In particular, for $N = 4$, the total spin $S = 1$, in agreement with all the previous publications except the DQMC model with parabolic confining potential by Pederiva, Umrigar and Lippurini (2000). For $N = 9$, we find a ground state with $S = 3/2$. However, we have shown (Matagne and Leburton submitted) that the spin sequence of the third shell depends on the splitting between its two degenerate d -states and its s -state. For $N = 13, 14, 15$ and 16 , $S = 1/2, 1, 1/2$ and 0 , respectively which confirms that the fourth shell has split into two sub-shells and the filling of the first sub-shell occurs in a way similar to the filling of the second shell. Thus, the filling of each sub-shell occurs in accordance to Hund's first rule.

Finally, Fig. 3 shows the electron density for $N = 6$ and $N = 12$ computed with LSDA. It is seen the electrons are delocalized which is confirmed by LSDA/DQMC. We disagree here with Yannouleas who, for the same densities, observes Wigner crystallization and spin density waves with a 2D HF model (Yannouleas and Landman 1999). We believe that this is due to the electron localization over-estimation, typical to the 2D models.

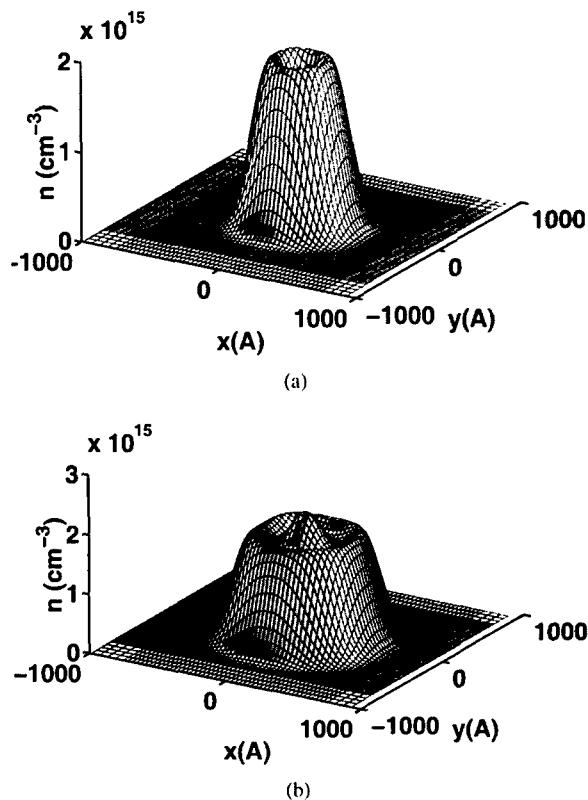


Figure 3. Electron densities for $N = 6$ (a) and $N = 12$ (b).

6. Conclusion

We have presented a new hybrid LSDA/DQMC model that includes the realistic 3D electron confining potential in the exact many-body calculation of electronic and spin configurations of VQDs. We have shown that, for QDs with high confinement, LSDA approximation is in excellent agreement with LSDA/DQMC. In addition, LSDA/DQMC confirms the atomic-like model of 2D shell structures obeying Hund's rule already predicted by LSDA.

Acknowledgments

This work is supported by the Beckman Institute, DARPA, under grant no. DAAD19-01-1-0659, and the Material Computer Center of the University of Illinois.

References

- Bolton F. 1994. Physical Review Letters 73(1): 158–161.
- Ezaki T., Mori N., and Hamaguchi C. 1997. Physical Review B 56: 6428–6431.

- Imamura K., Sugiyama Y., Nakata Y., Muto S., and Yokoyama N. 1995. Japanese Journal of Applied Physics 34: 1445.
- Lee I.-H., Rao V., Martin R.M., and Leburton J.-P. 1998. Physical Review B 57(15): 9035.
- Macucci M., Hess K., and Iafrate G.J. 1993. Physical Review B 48(23): 17354–17363.
- Matagne P. and Leburton J.-P. 2002. Physical Review B 65(23): 235323-1–712.
- Matagne P., Leburton J.-P., Destine J., and Cantraine G. 2000. Computer Modeling in Engineering and Sciences 1(1): 1–10.
- Pederiva F., Umrigar C.J., and Lippurini E. 2000. Physical Review B 62(12): 8120–8125.
- Rontani M., Rossi F., Manghi F., and Molinari E. 1999. Physical Review B 59(15): 10165–10174.
- Stopa M. 1996. Physical Review B 54(19): 13767–13783.
- Tarucha S., Austing D.G., Honda T., van der Hage R.J., and Kouwenhoven L.P. 1996. Physical Review Letters 77(17): 3613–3616.
- Wang Y. and Chou Y. 1993. Applied Physics Letters 63(16): 2257–2259.
- Wilkens T. 2001. Ph.D. Thesis, University of Illinois.
- Yannouleas C. and Landman U. 1999. Physical Review B 82(26): 5325–5328.



Theoretical Investigations of Spin Splittings and Optimization of the Rashba Coefficient in Asymmetric AlSb/InAs/GaSb Heterostructures

X. CARTOIXÀ

T.J. Watson, Sr., Laboratories of Applied Physics 128-95, California Institute of Technology,
Pasadena, CA 91125, USA

D.Z.-Y. TING

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

T.C. MCGILL

T.J. Watson, Sr., Laboratories of Applied Physics 128-95, California Institute of Technology,
Pasadena, CA 91125, USA

Abstract. We use the Effective Bond Orbital Model (EBOM) method to examine the spin splitting due to the Rashba effect in AlSb/InAs/GaSb asymmetric heterostructures. We have explored different thicknesses of the constituent materials and we have found for the resulting two-dimensional electron gas (2DEG) in the optimized structure a theoretical value of the Rashba constant $\alpha_R = 51 \times 10^{-10}$ eV·cm. This is, to our knowledge, the largest predicted value for this parameter. We provide an intuitive explanation for the behavior of the spins in the 2DEG. Finally, we study the effect of wide layers on the Rashba coefficient.

Keywords: two-dimensional electron gas, 2DEG, optimization, asymmetric quantum well, Rashba, EBOM

In recent years, interest in developing spin-sensitive devices (spintronics) (Wolf 2000, Gawel 2000, Das Sarma *et al.* 2000, Heinrich 2000) has fueled renewed investigations into spin phenomena in semiconductors. The aim is to control not only the spatial degrees of freedom of the electron, but also the spin degree of freedom. Useful spintronic devices can be devised if such control is achieved. A number of such devices have already been proposed (Datta and Das 1990, Fiederling *et al.* 1999, Morinaga and Shiiki 1999). Yet the search for phenomena which can lead to spin based devices is widespread (Schmidt *et al.* 2000, Malajovich *et al.* 2000, Awschalom and Kikkawa 1999).

We believe that the nearly lattice matched system comprised of InAs, GaSb and AlSb (McGill and Collins 1993, Miles *et al.* 1993) shows particular promise for spintronics. Studies of asymmetric quantum wells have been previously carried out both theoretically and

experimentally in other materials (Bychkov and Rashba 1984, Stein, Von Klitzing and Weimann 1983, Malcher, Lommer and Rossler 1986, Luo *et al.* 1988).

One of the phenomena that we expect to be of importance in the InAs/GaSb/AlSb heterojunction system is the Rashba effect (Bychkov and Rashba 1984). The precise contributions to the magnitude of the effect are a subject of some very recent studies (Silva, LaRocca and Bassani 1994, 1997). Previously, a number of studies have examined InAs quantum wells confined by InGaAs layers (Brosig *et al.* 1999, Matsuyama *et al.* 2000, Grundler 2000, Winkler 2000). In this study, we focus on the effects of the InAs/GaSb unique band offsets.

In this paper, we first perform a series of band structure calculations with the Effective Bond Orbital Model (EBOM) method (Chang 1988, Einevoll and Chang 1989), which takes the electron spin into account, for

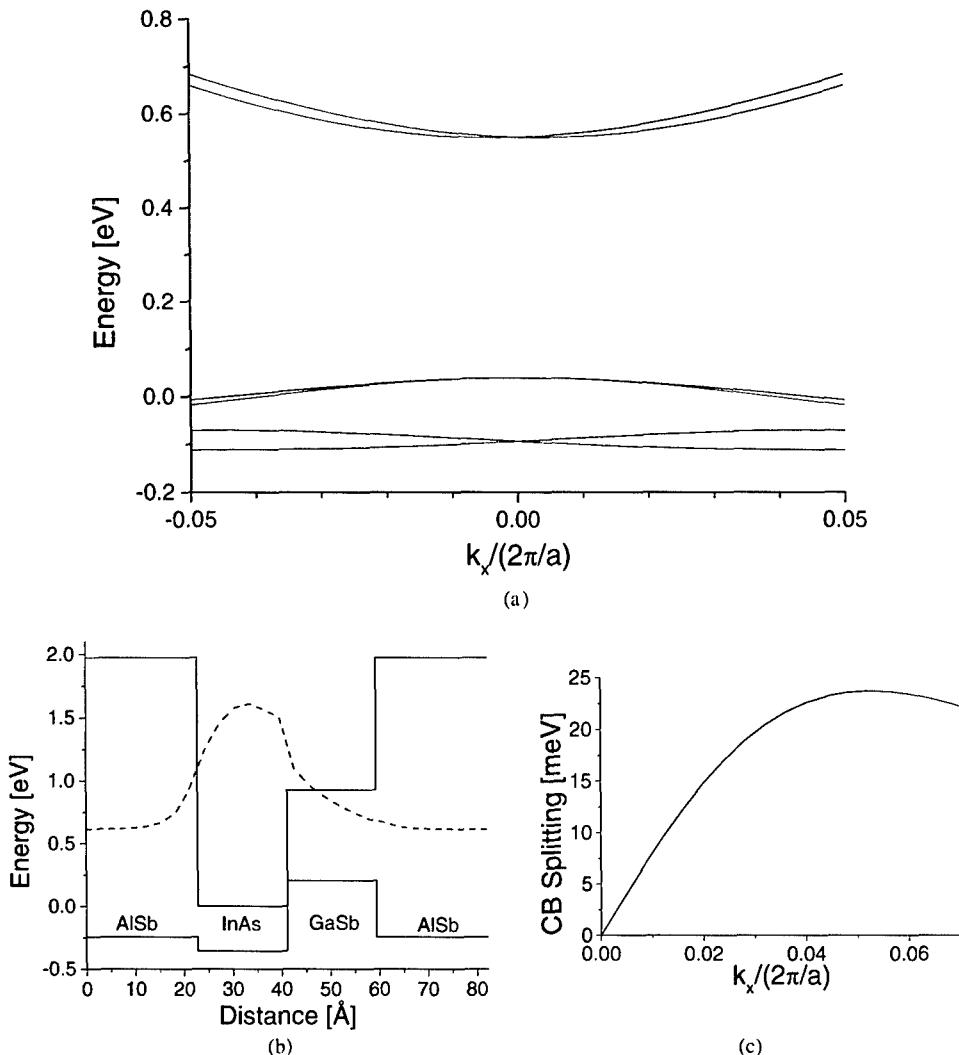


Figure 1. Plot (a) shows the calculated band structure of a 16 ML (monolayers) AlSb/6 ML InAs/6 ML GaSb superlattice near the Γ point for light holes, heavy holes and conduction electrons. Note the splitting in the states of the 2DEG, corresponding to the conduction band. The solid line in plot (b) shows the spatial arrangement of the band edges, while the dashed line corresponds to the wavefunction of a confined state. Plot (c) shows the splitting in the conduction band as a function of the electron wavevector along [100].

several AlSb/InAs/GaSb/AlSb quantum wells grown in the [001] direction. Later, we provide an intuitive physical picture of the processes inside the quantum well leading to the spin splitting. Finally, we find an optimized AlSb/InAs/GaSb structure yielding the largest Rashba coefficient and discuss a situation where it can be greatly diminished.

The EBOM method is a reformulation of Kane's eight-band $\mathbf{k} \cdot \mathbf{p}$ theory (Kane 1956) in the tight-binding framework. The EBOM technique incorporates realistic valence band structure. Hence, it is ideal for the calculation of the band structure of the two-dimen-

sional electron gas (2DEG), taking into account the strong coupling between the InAs conduction band and the GaSb valence band states (see plot b) in Fig. 1. It provides an accurate description of the band structure near the zone center for the conduction, heavy hole, light hole and split-off bands, including the spin-orbit splitting.

We use the standard approach of using a superlattice to find the band structure of the well, and we find that the superlattice cells are effectively decoupled when the AlSb thickness is of 8 monolayers (ML) or more, with a monolayer having a thickness of 3.048 Å.

When the inversion symmetry of a quantum well is broken, say by the introduction of a different layer at one of the well/barrier interfaces, a splitting in the energy levels in the conduction band appears (Bychkov and Rashba 1984). This splitting is linear in the electron momentum \mathbf{k} close to the Brillouin zone center.

Bychkov and Rashba derived the following Hamiltonian to describe this splitting in a two-band model (Bychkov and Rashba 1984)

$$H_{SO} = \alpha_R (\boldsymbol{\sigma} \times \mathbf{k}) \cdot \hat{z} \quad (1)$$

where α_R is the Rashba coefficient (dependent on the details of the heterostructure), $\boldsymbol{\sigma}$ are the Pauli spin matrices, \mathbf{k} is the crystal momentum of the electron in the 2DEG and \hat{z} is a unit vector parallel to the growth direction. When applied to quantum wells, the EBOM Hamiltonian yields the same results as Eq. (1) without any modification.

There are of course other symmetry breaking mechanisms for a well (Luo *et al.* 1988, Nitta *et al.* 1997) that will not be dealt with in this article. Our treatment does not account for the intrinsic inversion asymmetry existent in bulk zincblendes and the so called k^3 splitting originating from it (Dresselhaus 1955), either. This effect could lead to some corrections to our calculations, but we have found them to be on the order of a 20% at most (Cartoixà, Ting and McGill, unpublished).

We show in Fig. 1(a) the band structure of a 16ML (monolayers) AlSb/6ML InAs/6ML GaSb superlattice. We see that the two conduction subband energy levels are split. The hole states are also split, but their analysis in terms of spin angular momentum is much more complicated than in the electron case. The only point in \mathbf{k} space where the degeneracy remains is the Γ point. As can be deduced from Eq. (1), the band structure of the

2DEG possesses cylindrical symmetry to first order. As we can see in Fig. 1(c), the splitting of the conduction subbands is linear close to Γ , and from Eq. (1) it can be found to be

$$\Delta_R = 2\alpha_R k \quad (2)$$

where k is the magnitude of the electron wavevector. Figure 1(b) shows a plot of the band alignments of the composing materials and the probability distribution of an electron in the 2DEG.

The appearance of this splitting due to the lack of inversion symmetry can be intuitively understood in similar terms as the spin-orbit interaction. We choose our coordinate system such that the z axis points along the growth direction and the 2DEG lies on the $x-y$ plane. When the inversion symmetry is broken, there exists the possibility that a net electric field appears pointing in the z direction. In our case, the source of the net electric field would be the different band offsets in the AlSb/InAs and the InAs/GaSb confining interfaces. When the electrons move in the 2DEG plane, they will see in their comoving reference frame how a part of the original electric field has transformed into a magnetic field (Cohen-Tannoudji, Diu and Laloë 1977) due to the same relativistic effect leading to the spin-orbit interaction. One can see that this “effective” magnetic field (Johnson 1998) will be perpendicular to both the net electric field and the velocity—therefore, momentum—of the traveling electron. Finally, the electron spin couples to the in-plane magnetic field existing in its local frame in a Zeeman-like manner, giving rise to the split energy levels and a peculiar spin configuration in \mathbf{k} space. As seen in Fig. 2, the spins corresponding to a specific conduction subband will point in-plane either clockwise or counterclockwise in

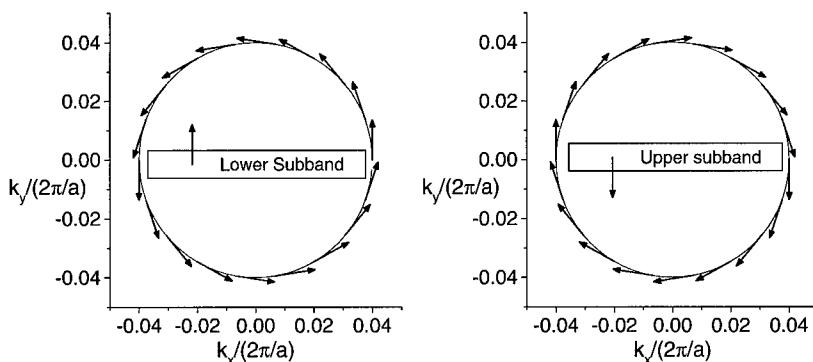


Figure 2. Calculated expectation value of the spin for states of a 2DEG lying on circles of constant k for the two conduction spin split subbands.

a circle of constant k magnitude in \mathbf{k} space, while the spins of the other subband point in the opposite way.

This spatial orientation of the spins in the conduction band could in principle be employed, with the help of a bias in the well plane, to partially magnetize the electrons in the well and generate a current in the 2DEG with a slight degree of spin polarization (Cartoixà *et al.* 2001). These effects will require a large Rashba coefficient if they are ever to be practically useful.

We have systematically explored structures with different layer thicknesses with the purpose of looking for the configuration optimizing the Rashba coefficient and delineate the role of each of the composing layers. The Rashba coefficient is calculated from the splitting at 0.2% of the zone edge along the [100] direction and then applying Eq. (2).

Figure 3 shows the Rashba coefficient as a function of the AlSb and GaSb thicknesses while keeping the InAs thickness fixed to 9 ML. We see that the Rashba coefficient is independent of the AlSb and GaSb thicknesses in a wide range. It only changes its value when the AlSb layer is very thin—the quantum well approximation ceases to be valid—or when the GaSb layer is less than 8 ML. In this case, a thinner GaSb layer reduces the amount of asymmetry in the quantum well and, therefore, reduces the Rashba coefficient. This behavior can be understood looking at the wavefunction in Fig. 1(b). We see the tail of the wavefunction vanishing into the AlSb on one side and into the GaSb on the

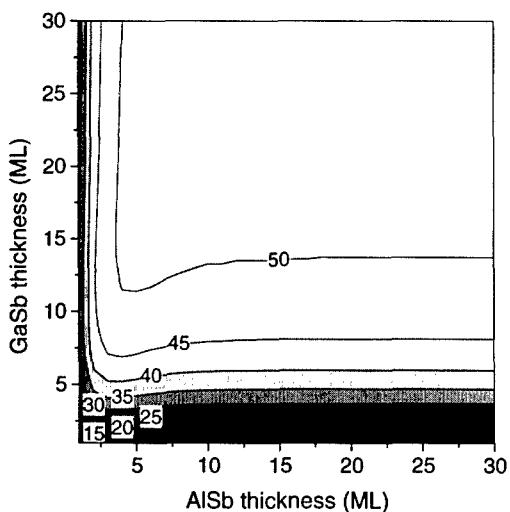


Figure 3. The Rashba coefficient as a function of AlSb and GaSb thicknesses for InAs thickness fixed to 9 ML. The numbers in the plot are contour lines of the Rashba coefficient in units of 10^{-10} eV·cm.

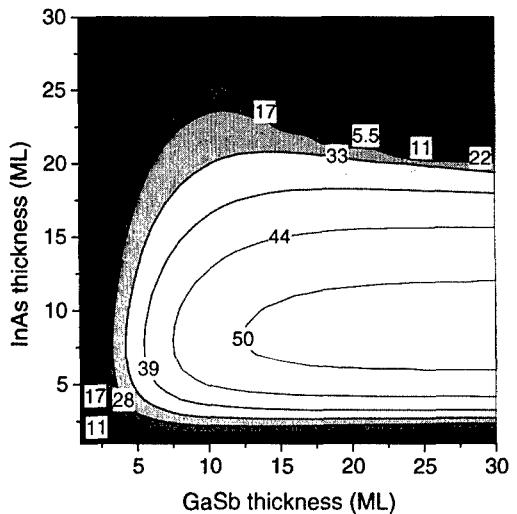


Figure 4. The Rashba coefficient as a function of InAs and GaSb thicknesses for AlSb thickness fixed to 16 ML.

other. Once the point where the tail of the wavefunction vanishes inside one layer has been reached; it is of little consequence that more monolayers of material are added, since that would affect a region where the electron is barely present. This explains the independence respect the AlSb and GaSb thicknesses for most of the range.

Figure 4 shows the Rashba coefficient as a function of the InAs and GaSb thicknesses while keeping the AlSb thickness fixed to 16 ML. We see that, in order to achieve a high Rashba coefficient, we need that the GaSb thickness be bigger than approximately 8 ML and the InAs layer be between 5 and 15 ML. We have seen that the AlSb thickness is of little importance as long as it is thick enough; therefore, these design rules apply for quantum wells with AlSb barriers thicker than 15 Å.

Another interesting feature in Fig. 4 is the region of thick GaSb and InAs where the Rashba coefficient is appreciably diminished—the dark region. This sudden reduction is due to the anticrossing of the InAs electron states with the GaSb hole states.

Figure 5 shows an example of such anticrossing. It displays the band structure of a 24 ML GaSb/24 ML InAs quantum well. In bulk, the InAs conduction band electron states lie below the GaSb valence band hole states (Yu, McCaldin and McGill 1992). For a thin quantum well, the InAs electron states are pushed up in energy, while the GaSb hole states are pushed down; thus making the well behave as if it were made of a direct bandgap material. However, for the case shown in Fig. 5, the GaSb and InAs layers are not thin enough for

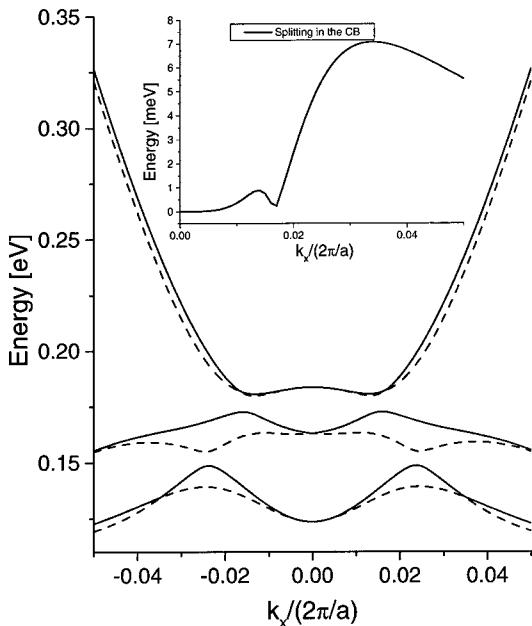


Figure 5. Band structure of a 12 ML AlSb/24 ML InAs/24 ML GaSb/12 ML AlSb quantum well, showing the anticrossing of the conduction and valence bands. The inset shows the splitting in the conduction band.

the valence and the conduction bands to be separated, and the anticrossing takes place.

We see in the inset in Fig. 5 that the amount of splitting only recovers its linear k dependence once the electron states don't couple with the hole states. In this situation, the Rashba coefficient defined as one half of the slope of the splitting at the Γ point loses much of its meaning, and it is more appropriate instead to look at the amount of splitting. We see, for example, that the splitting at $\mathbf{k} = (0.03, 0, 0) \text{ Å}^{-1}$ is reduced to about one third of the value for thin wells [cf. plot c in Fig. 1].

It is important to note that an eight band method is needed to fully consider the interplay of the conduction and valence bands leading to the reduction of the Rashba coefficient. Had we used a two band method, we would not have been able to detect this reduction.

In conclusion, we have shown that the AlSb/GaSb/InAs system shows promise to be the material system of choice to obtain large Rashba effect splittings. We have seen that the Rashba coefficient is critically dependent on the InAs thickness, while it shows almost no dependence on the AlSb and GaSb thicknesses once these are above some threshold value. We have determined the optimal thickness of the InAs layer. We have finally shown that, in order to achieve a large coefficient, it is important to design the sample in such a way that there

is no anticrossing between the electron states and the GaSb hole states.

Acknowledgment

We gratefully acknowledge the support of the Defense Advanced Research Projects Agency, as monitored by the Office of Naval Research under contract No. N00014-99-1-1006. A part of the work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Defense Advanced Research Projects Agency's Spins in Semiconductors (SpinS) program.

References

- Awschalom D.D. and Kikkawa J.M. 1999. Phys. Today 52: 33.
- Brosig S., Ensslin K., Warburton R., Nguyen C., Brar B., Thomas M., and Kroemer H. 1999. Phys. Rev. B 60: R13989.
- Bychkov Y.A. and Rashba E.I. 1984. JETP Lett. 39: 78.
- Cartoixa X., Ting D.Z.-Y., Daniel E.S., and McGill T.C. 2001 Superlatt. and Microstruct., 30(6). DOI: 10.1006/spmi.2002.1021.
- Cartoixa X., Ting D.Z.-Y., and McGill T.C. unpublished.
- Chang Y.C. 1988. Phys. Rev. B 37: 8215.
- Cohen-Tannoudji C., Diu B., and Laloë F. 1977. Quantum Mechanics, 2nd edn., Wiley, New York, p. 1215.
- Das Sarma S., Fabian J., Hu X.D., and Zutic I. 2000. Superlattice. Microst. 27: 289.
- Datta S. and Das B. 1990. Appl. Phys. Lett. 56: 665.
- Dresselhaus G. 1955. Phys. Rev. 100: 580.
- Einevoll G.T. and Chang Y.C. 1989. Phys. Rev. B 40: 9683.
- Fiederling R., Keim M., Reuscher G., Ossau W., Schmidt G., Waag A., and Molenkamp L.W. 1999. Nature 402: 787.
- Gawel R. 2000. Electron. Des. 48: 30.
- Grundler D. 2000. Phys. Rev. Lett. 84: 6074.
- Heinrich B. 2000. Can. J. Phys. 78: 161.
- Johnson M. 1998. Phys. Rev. B 58: 9635.
- Kane E.O. 1956. J. Phys. Chem. Solids 1: 82.
- Luo J., Munekata H., Fang F.F., and Stiles P.J. 1988. Phys. Rev. B 38: 10142.
- Malajovich I., Kikkawa J.M., Awschalom D.D., Berry J.J., and Samarth N. 2000. J. Appl. Phys. 87: 5073.
- Malcher F., Lommer G., and Rossler U. 1986. Superlatt. Microstruct. 2: 267.
- Matsuyama T., Kursten R., Meissner C., and Merkt U. 2000. Phys. Rev. B 61: 15588.
- McGill T.C. and Collins D.A. 1993. Semicond. Sci. Tech. 8: S1.
- Miles R.H., Schulman J.N., Chow D.H., and McGill T.C. 1993. Semicond. Sci. Tech. 8: S102.
- Morinaga A. and Shiiki K. 1999. Jpn. J. Appl. Phys. 38: 4741.
- Nitta J., Akazaki T., Takayanagi H., and Enoki T. 1997. Phys. Rev. Lett. 78: 1335.
- Schmidt G., Fiederling R., Keim M., Reuscher G., Gruber T., Ossau W., Waag A., and Molenkamp L.W. 2000. Superlattice. Microst. 27: 297.

- Silva E.A.D.E., LaRocca G.C., and Bassani F. 1994. Phys. Rev. B 50: 8523.
- Silva E.A.D.E., LaRocca G.C., and Bassani F. 1997. Phys. Rev. B 55: 16293.
- Stein D., Von Klitzing K., and Weimann G. 1983. Phys. Rev. Lett. 51: 130.
- Winkler R. 2000. Phys. Rev. B 62: 4245.
- Wolf S.A., Supercond J. 2000. 13: 195.
- Yu E.T., McCaldin J.O., and McGill T.C. 1992. Solid State Phys. 46: 1.



Modeling Spin-Dependent Transport in InAs/GaSb/AlSb Resonant Tunneling Structures

D.Z.-Y. TING

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

X. CARTOIXÀ AND T.C. MCGILL

Department of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA

D.L. SMITH

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

J.N. SCHULMAN

HRL Laboratories, LLC, Malibu, CA 90265, USA

Abstract. The Rashba effect resonant tunneling diode is a candidate for achieving spin polarizing under zero magnetic field using only conventional non-magnetic III–V semiconductor heterostructures. We point out the challenges involved based on simple arguments, and offer strategies for overcoming these difficulties. We present modeling results that demonstrate the benefits of the InAs/GaSb/AlSb-based asymmetric resonant interband tunneling diode (a-RITD) for spin filtering applications.

Keywords: resonant tunneling, interband tunneling, Rashba effect, spin filter

1. Introduction

Recent theoretical studies suggest the possibility of polarizing electron spins by resonant tunneling (de Andrada e Silva and La Rocca 1999), and obtaining spin-polarized current in resonant tunneling heterostructures at zero magnetic field (Voskoboinikov *et al.* 2000) using conventional non-magnetic semiconductor heterostructures. A typical resonant tunneling spin-filtering device structure consists of double barriers surrounding an asymmetric quantum well, where quantized states are spin-split by the Rashba effect (Bychkov and Rashba 1984). In this work we report our theoretical analysis of spin polarization effects in InAs/GaSb/AlSb resonant tunneling structures. Heterostructures of InAs/GaSb/AlSb are strong candidates for pronounced Rashba spin splitting because of the large spin-orbit interaction in InAs and GaSb, and

the availability of both InAs and GaSb for the construction of highly asymmetric quantum wells. The non-common anion InAs/GaSb and InAs/AlSb heterointerfaces also present opportunities for engineering interface potentials for optimizing Rashba spin splitting. Indeed, our calculations show that it is possible to obtain rather large Rashba spin splitting in AlSb/InAs/GaSb superlattices even in the absence of a transverse electric field (Cartoixà, Ting and McGill in press, 2002).

The InAs/GaSb/AlSb material system allows for a variety (type-I, type-II staggered, and type-II broken-gap) of band alignments. In addition to conventional intraband resonant tunneling structures, the availability of the type-II broken-gap band offset also allows us to fabricate resonant interband tunnel diodes (RITDs), where the quasibound states have opposite k -parallel dispersions to those in the electrodes. In the following

sections we will discuss spin-dependent tunneling in both the intraband and the interband resonant tunneling structures. In particular, we will highlight salient features of the RITD for applications to spin filtering.

2. Rashba Effect Resonant Tunneling Spin Filter

Spin splitting can result from the lifting of Kramers degeneracy through the removal inversion symmetry. Bulk zinc-blende or wurzite compound semiconductors exhibit microscopic inversion asymmetry in their lattice structures. Dresselhaus (1955) has shown that this bulk inversion asymmetry (BIA) leads to a spin-orbit induced splitting whose magnitude is proportional to k^3 for small k , where k is the electron wave number. The Dresselhaus mechanism can also lead to an additional splitting linear in k in 2D systems (Eppenga and Schuurmans 1988). Rashba showed that in heterostructures, structural inversion asymmetry (SIA) from interface fields can lead to a spin splitting which is also linear in k (Bychkov and Rashba 1984). It is believed that in InAs/GaSb and InAs/AlSb systems, spin splitting is primarily due to the Rashba rather than the Dresselhaus effect (Luo *et al.* 1990, Chen *et al.* 1993). Therefore in this work we include only effects due to SIA; the treatment of BIA will be reported elsewhere (Cartoixà, Ting and McGill unpublished). Figure 1 shows the band structure of an InAs/GaSb/AlSb superlattice structure with composite InAs/GaSb wells calculated using the effective bond orbital model (Chang 1988). The structural inversion asymmetry induces spin splitting in both conduction and valence subband states (localized, respectively in InAs and GaSb layers). Spin splitting in the lowest conduction subband exceeds 20 meV for in-plane wave vector $k_{\parallel} = 0.05(2\pi/a)$; splitting for all subbands vanish at the zone center.

The Rashba effect may be exploited in conjunction with resonant tunneling for spin filtering using, for example, a double barrier heterostructure (DBH) containing an asymmetric quantum well. In our case, we use an asymmetric composite InAs/GaSb well, surrounded by AlSb barriers and InAs electrodes. When designed with appropriate layer thicknesses, this device can be made to operate either, under low bias, in the resonant interband tunneling (Söderström, Chow and McGill 1989) regime where electrons traverse valence subband states in GaSb, or, under moderate bias, in the intraband resonant tunneling regime where electrons traverse conduction subband states in InAs. The quasibound states in this structure are spin-split due to the Rashba

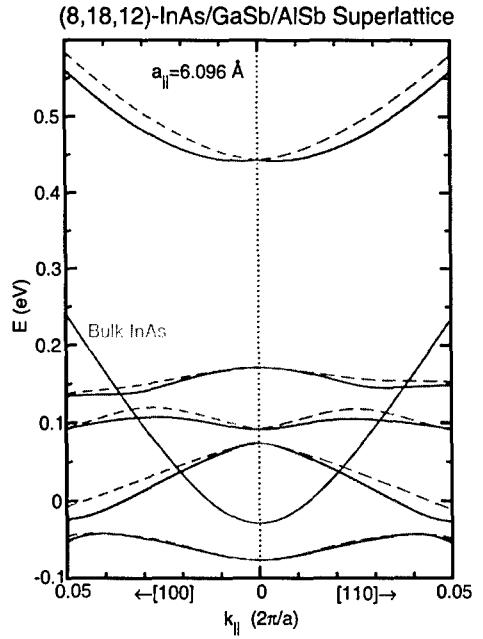


Figure 1. Band structure of a superlattice, where each period contains 8 ML (monolayers) of InAs, 18 ML of GaSb, and 12 ML of AlSb. The unstrained band edges of $E_c^{InAs} = 0$ and $E_v^{GaSb} = 0.204$ eV are used. The substrate lattice constant is taken to be that of GaSb. The lowest conduction band of InAs, strained to the GaSb lattice constant in the in-plane directions, is also shown for reference.

effect. Large spin splitting is desirable for spin filtering applications. Near the zone center Rashba spin splitting is linear in k , and is given by $\Delta E = 2\alpha|k|$, where α is known as the Rashba coefficient. We compute α associated with the lowest conduction subband for our resonant tunneling structure, and show in Fig. 2 how it can be modified by an applied E field. For comparison, we also show the results for a symmetric structure that shows no spin splitting under zero bias. Evidently, the applied E field alters the degree of inversion asymmetry and hence the Rashba spin splitting.

The spin of a resonantly transmitted electron is aligned with that of the quasibound state traversed (de Andrade e Silva and La Rocca 1999). Figure 3 shows the spin-dependent transmission coefficient spectra calculated using the multiband quantum transmitting boundary method (MQTBM) (Ting, Yu and McGill 1992) for the asymmetric DBH described above. Incident electrons with $+z$ and $-z$ polarizations are used to probe quasibound states with k_{\parallel} along the x direction, which have $+y$ and $-y$ spin polarizations (see discussions below). Note that for energies above or below the resonances, the transmitted electron retains the spin direction of the incident electron. For energies in

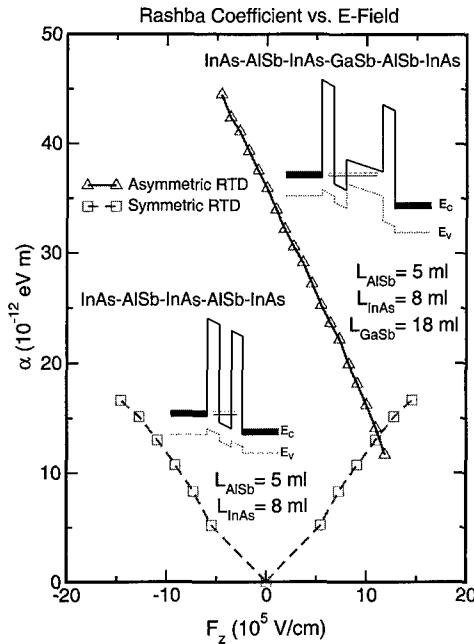


Figure 2. Dependence of Rashba coefficients on applied electric field for a symmetric and an asymmetric resonant tunneling structure. Band diagrams are shown as insets.

between the resonances, spins are flipped during the tunneling process. On resonance, however, transmitted states are spin polarized by the quasibound states, independent of the incident spin direction. Since only on-resonance transmission probabilities are significant, the spin polarization of transmitted current is primarily determined by quasibound state properties, which we examine below.

3. Challenges and Strategies for Designing a Rashba Effect Spin Filter

To demonstrate the challenges involved in designing a resonant tunneling spin filter based on the Rashba effect, we examine the spin directions of the quasibound states. In general, spin-orbit interaction is given by $H_{\text{so}} = [h/(2mc)^2]\boldsymbol{\sigma} \cdot \nabla V \times \mathbf{p} \equiv (g/2)\mu_B \boldsymbol{\sigma} \cdot \mathbf{B}_{\text{eff}}$. Spins of quantum well quasibound states align with the effective magnetic field \mathbf{B}_{eff} , and from the form of \mathbf{B}_{eff} we readily conclude: (1) $\mathbf{B}_{\text{eff}} \perp \nabla V$. Since we consider only SIA, spatial variations of V are along the growth direction, implying that spins are in the plane of the quantum well. (2) $\mathbf{B}_{\text{eff}} \perp \mathbf{p}$, or, since \mathbf{k}_{\parallel} is a good quantum number, $\mathbf{B}_{\text{eff}} \perp \mathbf{k}_{\parallel}$. Hence spins are perpendicular to \mathbf{k}_{\parallel} . (3) $|\mathbf{B}_{\text{eff}}| \propto |\mathbf{k}_{\parallel}|$. Thus spin splitting vanishes at the zone center.

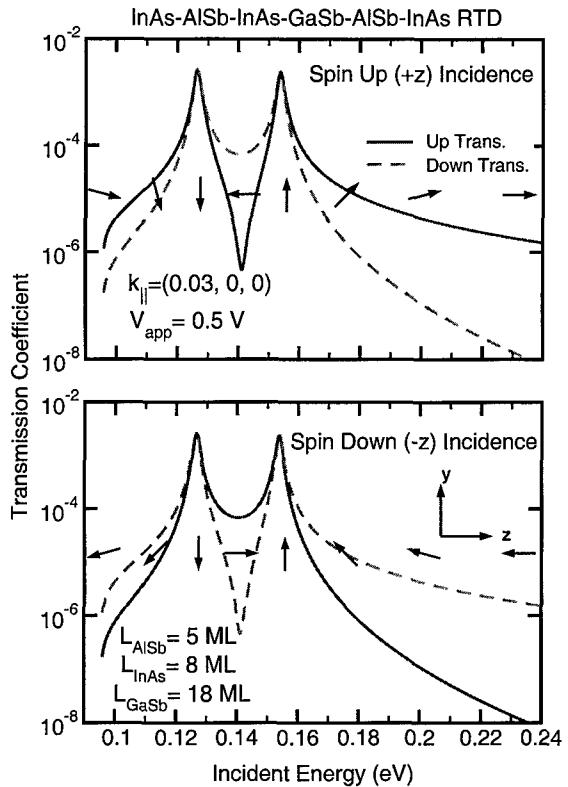


Figure 3. Spin-dependent transmission coefficient spectra of a double barrier structure with an asymmetric composite InAs-GaSb well, AlSb barriers, and InAs electrodes. A bias of 0.5 V is applied over the active region. Resonant tunneling through conduction band quasibound states are examined. The in-plane wave vector is $\mathbf{k}_{\parallel} = (0.03, 0, 0)(2\pi/a)$. Spin directions are defined with respect to the growth direction, taken as the z -axis. Top and bottom panels show results for incident electrons with $+z$ and $-z$ spin polarizations, respectively. Transmission probabilities into states with both $+z$ and $-z$ polarizations are shown in each panel as solid and dashed lines, respectively. Overlay arrows are used to indicate the spin directions of the transmitted states.

The analysis above reveals the difficulties involved in using the quasibound states in resonant tunneling structures for spin alignment. First, at any given \mathbf{k}_{\parallel} , the two spin-split states have opposite spins. While this is exactly the property we wish to exploit for spin filtering, we also need to ensure that we could resolve the spin split states so we can preferentially select one of the spin polarizations. The strategy for achieving this is to maximize spin splitting, and use resonant tunneling to resolve the states. There is theoretical evidence that large Rashba coefficient may be obtained in the InAs/GaSb/AlSb systems (Cartoixa, Ting and McGill in press) Next, the $+\mathbf{k}_{\parallel}$ and $-\mathbf{k}_{\parallel}$ states within a given

spin-split subband have opposite spins. In a typical resonant tunneling diode, incident electrons come from a reservoir in thermal equilibrium, occupying $+k_{\parallel}$ and $-k_{\parallel}$ states with equal probability. Thus the ensemble of transmitted electrons yields no net spin polarization. To address this issue, VoskoBoynikov *et al.* (2000) proposed the application of a small lateral (perpendicular to the growth direction) E -field in the source region of the resonant tunneling diode to shift the incident electron distribution towards, say, the positive k_x side in k -space. The resonantly transmitted currents originating from this non-equilibrium distribution would then show spin polarization. Finally, since spin splitting is linear in k near (and vanishes at) the zone center, resolving the spin-split states there is not feasible. In the next section we discuss how the interband tunneling mechanism might be used to address this issue.

4. Asymmetric Resonant Interband Tunneling Diodes

The resonant interband tunneling condition (Söderström, Chow and McGill 1989) is illustrated in Fig. 1. It shows that a number of valence subband states in the asymmetric quantum well are above the InAs conduction band edge. Thus in our double barrier structure, conduction band electrons can tunnel from one InAs electrodes to the other through valence subband states under low bias. Figure 4 shows transmission coefficient spectra in interband as well as intraband regimes. For this calculation, we intentionally align the incident electron spins according to the spin directions of the resonances. Thus each of the incident spin polarization only couples to one of the spin-split resonances. The top panel shows that in the intraband tunneling case the resonant transmission probability through the two spin-split lowest conduction subband states (cb1) are approximately equal. On the other hand, in the interband tunneling case shown on the bottom panel, transmission probability through the highest heavy hole ($hh1$) states is much higher for the $+y$ than the $-y$ spin polarization. The transmission peak strength ($T_{\max} \Delta E$, peak height times peak width) of the $+y$ channel is approximately 17 times larger than that of the $-y$ channel.

Figure 5 summarizes spin-dependent resonant interband tunneling properties of our structure. We focus on the $hh1$ result since we intend to use it for spin filtering; the $lh1$ states can be pushed away from $hh1$ states by changing layer widths and composition of the well, and

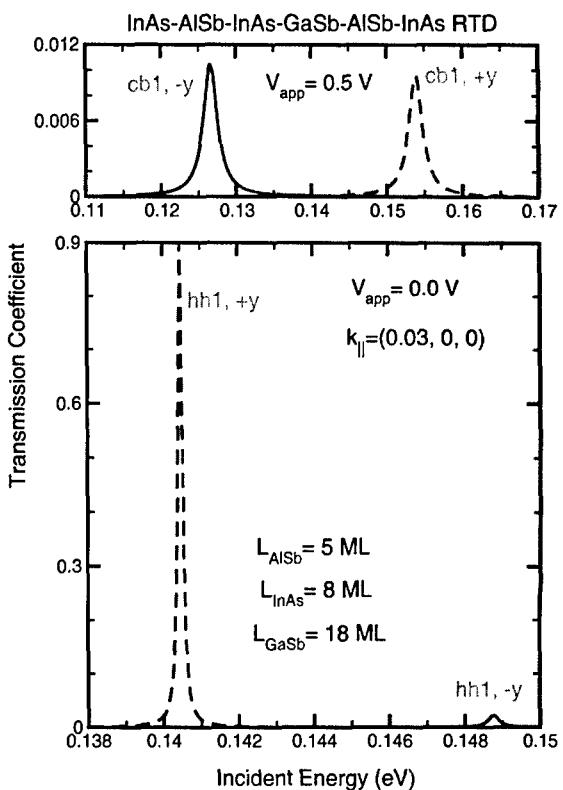


Figure 4. Transmission coefficient spectra for an InAs/GaSb/AlSb double barrier structure with an asymmetric quantum well. The top and bottom panels respectively show results for intraband and interband tunneling regimes, reached under different biasing conditions. The in-plane wave vector is $k_{\parallel} = (0.03, 0, 0)(2\pi/a)$. The dashed and solid lines represent results for incident electron with $+y$ and $-y$ spin polarizations, respectively.

also lowered by biasing. The $hh1$ states have a number of attractive features for spin filtering application. The $hh1$ energies decrease with increasing k_{\parallel} , allowing the selection of states with k_{\parallel} away from the zone center by setting the Fermi level in the incoming electrode to be below the energy of the zone center $hh1$ states. Also, $hh1$ peak strengths are exceedingly weak near the zone center due to inadequate hole mixing (Ting, Yu and McGill 1992). This also allows us to filter out zone-center states. Finally, the pronounced difference in the strengths of the two $hh1$ spin channels can be exploited for spin filtering.

5. Summary and Discussions

We discuss the basic principles of the Rashba effect resonant tunneling spin filter. We point out the chal-

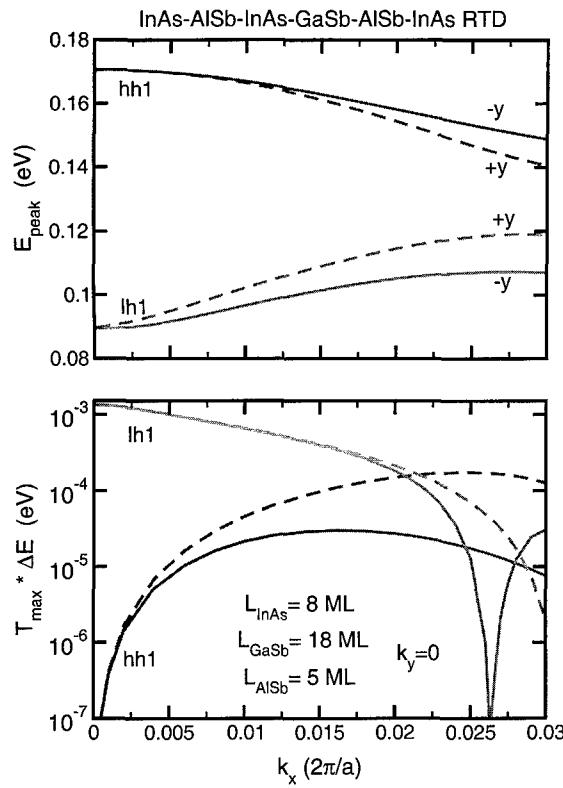


Figure 5. Resonant transmission peak position (top panel) and strength (bottom panel) as functions of in-plane wave vector k_x for the highest heavy hole (hh1) and light hole (lh1) states. The dashed and solid lines represent results for incident electron with $+y$ and $-y$ spin polarizations, respectively.

lenges based on quite general arguments, and offer strategies for overcoming these difficulties. In particular, we present modeling results, which demonstrate the advantages of using the InAs/GaSb/AISb-based asymmetric resonant interband tunneling diode (a-RITD).

The a-RITD can effectively exclude tunneling through states near $\mathbf{k}_{||} = 0$ where Rashba spin splitting vanishes, and spin selectivity is difficult. Away from the zone center, a-RITD can provide strong spin selectivity. When coupled with an emitter capable of k -space selectivity, which is a challenge in itself, the a-RITD should be able to achieve spin filtering in semiconductors under zero magnetic field using only conventional non-magnetic III-V semiconductor heterostructures.

Acknowledgments

The authors thank D.H. Chow and T.F. Boggess for helpful discussions. This work was supported by the Defense Advanced Research Projects Agency (DARPA) Spins in Semiconductors (SpinS) program.

References

- Bychkov Y.A. and Rashba E.I. 1984. JETP Lett. 39: 78.
- Cartoixà X., Ting D.Z.-Y., and McGill T.C. Journal of Computational Electronics, in press.
- Cartoixà X., Ting D.Z.-Y., and McGill T.C. unpublished.
- Chang Y.C. 1988. Phys. Rev. B 37: 8215.
- Chen G.L., Han J., Huang T.T., Datta S., and James D.B. 1993. Phys. Rev. B 47: 4084.
- de Andrade e Silva E.A. and La Rocca G.C. 1999. Phys. Rev. B 59: 15583.
- Dresselhaus G. 1955. Phys. Rev. 100: 580.
- Eppenga R. and Schuurmans M.F.H. 1988. Phys. Rev. B 37: 10923.
- Luo J., Munekata H., Fang F.F., and Stiles P.J. 1990. Phys. Rev. B 41: 7685.
- Söderström J.R., Chow D.H., and McGill T.C. 1989. Appl. Phys. Lett. 55: 1094.
- Ting D.Z.-Y., Yu E.T., and McGill T.C. 1992. Phys. Rev. B 45: 3583.
- Voskoboynikov A., Lin S.S., Lee C.P., and Tretyak O. 2000. J. Appl. Phys. 87: 387.



Tunneling through Thin Oxides—New Insights from Microscopic Calculations

M. STÄDELE

Infineon Technologies, CPR ND, Otto-Hahn-Ring 6, D-81730 Munich, Germany

B. TUTTLE

Department of Physics, Penn State University, Erie, Pennsylvania 16563-0203, USA

B. FISCHER AND K. HESS

Beckman Institute, University of Illinois, Urbana, IL 61801, USA

Abstract. In this paper, we summarize our recent efforts to analyze transmission probabilities of extremely thin SiO_2 gate oxides using *microscopic models* of $\text{Si}[100]\text{-SiO}_2\text{-Si}[100]$ heterojunctions. We predict energy-dependent tunneling masses and their influence on transmission coefficients, discuss tunneling probabilities and analyze effects arising from the violation of parallel momentum conservation. As an application of the present method, gate currents in short bulk MOSFETs are calculated, including elastic defect-assisted contributions.

1. Introduction

Tunneling currents through a few atomic layers ($\approx 1 \text{ nm}$) thin gate oxides represent one of the major factors that may soon limit the gigascale integration of ultrasmall metal-oxide-semiconductor field effect transistors (MOSFETs) (ITRS 1999). It is obvious that for such thin layers the microscopic structure of the oxide and its interface with Si influences tunneling currents drastically. Accordingly, simple and widely used models for calculating oxide transmission probabilities such as the Wentzel-Kramers-Brillouin approach (Duke 1969) or the effective-mass based multiple scattering theory (Ando and Itoh 1987) become more and more questionable as the oxides are scaled down.

To overcome fundamental limitations, we have calculated transmission probabilities and gate leakage currents for microscopic oxide models that were constructed using first-principles density-functional methods. Transmission coefficients were subsequently calculated using a tight-binding formalism and combined with Monte Carlo device simulation data. Among other

issues, such an approach allows one to estimate the influence of bond distortions, interface structure, and resonant tunneling through defects on the transmission, to predict the intrinsic decay properties of the states within the oxide band gap, to assess the degree to which a bulk band structure picture can help in understanding tunneling through very thin oxides, to investigate effects due to violation of \mathbf{k}_{\parallel} conservation in transmission and reflection, and to assess the validity of effective-mass based approaches. In this paper, we will briefly discuss some of the most important results, referring the interested reader to a more extended discussion in Städele, Tuttle and Hess (2001) and Städele *et al.* (to appear).

2. Microscopic Calculation of Oxide Transmission Coefficients

In this section, we briefly summarize the computational procedure that we have utilized to obtain tunneling probabilities. Our strategy consists of two steps: construction of the microscopic models

and calculation of the corresponding transmission coefficients.

The microscopic supercell models of Si[100]-SiO₂-Si[100] heterojunctions we have used were constructed by sandwiching unit cells of the tridymite or β -quartz polytype of SiO₂ between two Si[100] surfaces. Subsequently, both the coordinates of the atoms and the supercell lengths perpendicular to the interface were relaxed using gradient-corrected (GGA) local-density calculations (Dreizler and Gross 1990). In the following, these supercells will be also referred to as $n \times n$ to indicate that the lateral dimension is some multiple of the periodicity of the silicon surface (which corresponds to a 1×1 cell). Figure 1 shows a ball-and-stick skeleton of the tridymite-based cell as an example.

Reflection and transmission coefficients of the supercells described above were calculated using a transfer-matrix type scheme embedded in a tight-binding framework (Städle, Tuttle and Hess 2001, Strahberger 1999, Strahberger and Vogl 1999). We solve the Schrödinger equation with open boundary conditions for the whole junction at a fixed energy E (measured relative to the Si conduction band minimum on the channel side of the oxide) and in-plane momentum $\mathbf{k}_{\parallel}^{n \times n}$ (that is a good quantum number due to the lateral periodicity) in a 'layer-orbital basis' comprised of the following states:

$$|\sigma, \mathbf{k}_{\parallel}^{n \times n}\rangle = \frac{1}{\sqrt{N_{\parallel}^{n \times n}}} \sum_{\mathbf{R}_{\parallel}^{n \times n}} e^{i\mathbf{k}_{\parallel}^{n \times n}\mathbf{R}_{\parallel}^{n \times n}} |\sigma, \mathbf{R}_{\parallel}^{n \times n}\rangle. \quad (1)$$

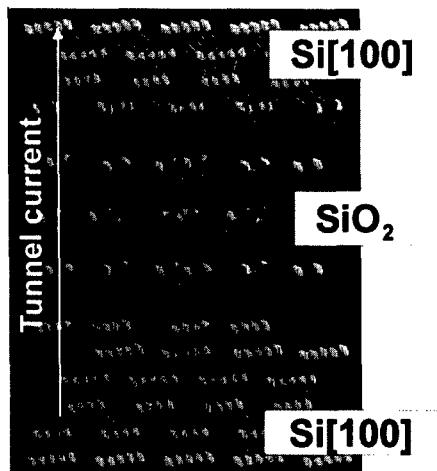


Figure 1. Model of a Si[100]/SiO₂/Si[100] tridymite heterojunction with an ultrathin (1.3 nm) gate oxide region. The darker (lighter) balls denote O (Si) atoms.

Here, all the orbitals in a layer (=any collection of atoms in the cell) are lumped into the index σ , $\mathbf{R}_{\parallel}^{n \times n}$ designates an in-plane Bravais lattice vector of the $n \times n$ structure, $|\sigma, \mathbf{R}_{\parallel}^{n \times n}\rangle$ is a particular localized orbital, and $N_{\parallel}^{n \times n}$ the number of unit cells per layer. A state propagating towards the oxide from the channel side of the junction, characterized by $E, \mathbf{k}_{\parallel}^{n \times n}$ and its wavevector component $k_{\perp,i}^{in}$ normal to the interface, is scattered into sets of reflected and transmitted states (characterized by wavevector components $k_{\perp,j}^{out}$). From the scattering wavefunctions, transmission amplitudes $t_{\mathbf{k}_{\parallel}}(k_{\perp,i}^{in} \rightarrow k_{\perp,j}^{out})$ and dimensionless transmission coefficients

$$T_I(E) = \frac{1}{A_{\mathbf{k}_{\parallel}^{n \times n}}} \int d\mathbf{k}_{\parallel}^{n \times n} T(E, \mathbf{k}_{\parallel}^{n \times n}) \quad (2)$$

$$= \frac{1}{A_{\mathbf{k}_{\parallel}^{n \times n}}} \int d\mathbf{k}_{\parallel}^{n \times n} \times \sum_i \sum_j |t_{\mathbf{k}_{\parallel}}(k_{\perp,i}^{in} \rightarrow k_{\perp,j}^{out})|^2 \frac{v_{\perp,j}^{out}}{v_{\perp,i}^{in}} \quad (3)$$

are obtained. Here, $A_{\mathbf{k}_{\parallel}^{n \times n}}$ is the area of the planar Brillouin zone of the junction, and $v_{\perp,i}^{in}$ and $v_{\perp,j}^{out}$ denote the components of the group velocities of the incident and transmitted Si bulk states (with wavevector components $k_{\perp,i}^{in}$ and $k_{\perp,j}^{out}$ normal to the interfaces). An sp^3 TB basis with second-nearest neighbor interactions for both silicon (Grosso and Piermarocchi 1995) and the oxide (Städle, Tuttle and Hess 2001) was used. The Si TB conduction band structures agree fairly well with experiment for energies up to 3 eV (Grosso and Piermarocchi 1995). The oxide parameters were chosen to yield a band gap of 8.95 eV and to reproduce the GGA effective masses of the lowest SiO₂ conduction band in the [100] direction ($0.42 m_0$ and $0.6 m_0$ for the tridymite and β -quartz structures, respectively).

3. Results and Discussion

3.1. Energy-Dependent Oxide Tunneling Masses

The present microscopic models allow one to predict the intrinsic decay properties of the wavefunctions in SiO₂. In most practical calculations, the decay constants are fixed implicitly by choosing an (possibly energy-dependent) effective mass at the energetically nearest band extremum. In particular, Franz-type (Franz 1956, Khairurrijal *et al.* 1999, Av-Ron *et al.* 1981, Brar, Wilk and Seabaugh 1996, Maserjian 1974,

Maserjian and Zaman 1982, Krieger and Swanson 1981) or $\mathbf{k} \cdot \mathbf{p}$ -type (Zhakarova, Ryshii and Pesotzkii 1994) dispersions have been used previously with considerable success but little justification. We have analyzed the complex bands of the present oxide models and find that (i) only one single complex band is relevant for electron tunneling, (ii) several different bands are involved in hole tunneling, and (iii) all complex oxide bands are highly nonparabolic. For electrons, the nonparabolicity can be cast into an energy-dependent tunneling mass via the equation $\hbar \Im k_{\perp}(\varepsilon) = \sqrt{2m_t(\varepsilon)\varepsilon}$. Here, $\Im k_{\perp}$ is the smallest imaginary part of the complex \mathbf{k} vectors in the oxide gap, $\varepsilon := E_{CBO} - E$ measures the energy from the conduction band minimum of SiO_2 toward the valence region. For holes, a tunneling mass can be derived analogously. Electron and hole masses determined from the β -quartz and tridymite models are shown in Fig. 2. It is apparent that the parabolic approximation ($m/m_{band\ edge} = 1$) fails completely for both electrons and holes. Furthermore, we note that the electron masses for both structural models almost coincide when normalized by the mass at the bottom of the lowest conduction band. The energy dependence of the electron mass shown in Fig. 2 might therefore be a more general feature of electron tunneling in SiO_2 . At the top of the valence band, the hole masses are very large ($\approx 3 m_0$ for the tridymite and $\approx 16 m_0$ for the β -quartz model). However, this is largely canceled by the strong nonparabolicity of the

complex hole bands which leads for both models to an $\Im k_{\perp}$ on the order of $\approx 0.5 \text{ \AA}^{-1}$ a few eV above the valence band maximum for both models (see inset of Fig. 2).

3.2. Energy Dependence of Transmission

The energy dependence of the integrated transmission $T_I(E)$ (this quantity is relevant for the calculation of currents) is shown in Fig. 3, which also includes effective-mass based results¹ with a parabolic and the energy-dependent electron mass from Fig. 2. Due to averaging effects, the integrated tight-binding transmission is much smoother than the individual coefficients $T(E, \mathbf{k}_{\parallel})$, which can change abruptly by 1–2 orders of magnitude when new bands of different symmetry appear (not shown). The parabolic effective-mass approximation overestimates $T_I(E)$ for oxides thicknesses t_{ox} smaller than $\approx 1 \text{ nm}$, by up to two orders of magnitude. As t_{ox} increases, the tight-binding transmission $T_I(E)$ is underestimated at low energies and overestimated at higher energies. The higher slope of the transmission obtained in the parabolic effective-mass approximation is consistent with the findings for the tunneling masses: as E increases, the overestimate of m_t and of $\Im k_{\perp}$ by the effective-mass calculation decreases, leading to a relative increase of the effective-mass transmission. Using the correct tight-binding dispersion of

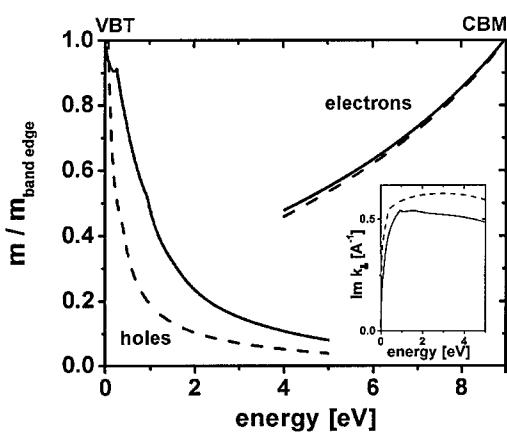


Figure 2. Effective energy-dependent tunneling masses in the SiO_2 band gap for electrons and holes, given in units of the mass at the nearest band edge (valence band top (VBT) for holes, conduction band minimum (CBM) for electrons). Solid and dashed lines refer to the tridymite and β -quartz models, respectively. The inset shows the energy dependence of the smallest imaginary part of the complex wavevectors for both models.

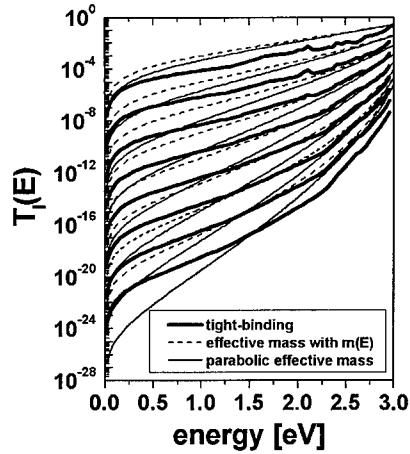


Figure 3. Tight-binding transmission coefficients versus energy, $T_I(E)$, of tridymite-type oxide models with thicknesses of 0.73, 1.28, 1.83, 2.38, 2.93, 3.48, 4.03, and 4.58 nm (thick solid lines). Also shown are effective-mass data obtained using an energy-independent tunneling mass (thin solid lines) and the energy-dependent tunneling mass given in Fig. 2 (thin dashed lines). Zero oxide bias has been assumed.

the imaginary bands (dashed lines in Fig. 2) in an effective-mass calculation leads to qualitatively correct slopes for $T_I(E)$; however, the absolute values are typically overestimated by one to two orders of magnitude, almost independent t_{ox} and of E . For the β -quartz model, we have obtained quantitatively very similar results.

A possible reason for much of this discrepancy may be that the 1D effective-mass based transmission calculation underestimates the 3D band structure mismatch of Si and SiO_2 and therefore the reflection from the Si/ SiO_2 interface; due to the weak sensitivity to barrier thickness, only a small part of this effect stems from the SiO_2 regions. In addition, differences of 1–2 orders of magnitude are also observed for the transmission coefficients of electrons with energies sufficiently high that almost no tunnel barrier exists (compare Fig. 3 at $E = 3$ eV).

These findings have an important implication: in a model for gate leakage currents based on effective-mass calculations, a fitting-parameter adjustment has to compensate for the overestimated transmission coefficients. If a Franz-type tunneling mass (which is close to the mass obtained from the present tight-binding calculations) were employed, such an adjustment would be an overestimate of t_{ox} (by about 0.3 nm) or of the carrier density at the Si/ SiO_2 interface. In the case of a parabolic, energy-independent mass, the picture would be more complicated: for very thin oxides, the difference between the effective-mass and tight-binding transmission coefficients is positive and changes sign as the thickness increases. Therefore, the slope of a curve that shows the measured t_{ox} versus the fitted t_{ox} is expected to be greater than one (we assume implicitly that the measured thickness is close to the thickness of our models). An indication of this is found when comparing effective-mass results with XPS thickness measurements (see Fukuda *et al.* (1998)).

3.3. Violation of Parallel Momentum Conservation in Transmission and Reflection

For the tridymite model, we have investigated coupling effects between states having the same $\mathbf{k}_{\parallel}^{n \times n}$ but different bulk \mathbf{k}_{\parallel} . Interestingly, our main conclusions turn out to be virtually independent of the choice of $\mathbf{k}_{\parallel}^{n \times n}$, the energy, or the applied bias voltage. Therefore, we illustrate the main effects for the case of the tridymite model at $E = 1.5$ eV and a $\mathbf{k}_{\parallel}^{n \times n}$ vector that allows for the coupling of 10 states with (generally different) bulk

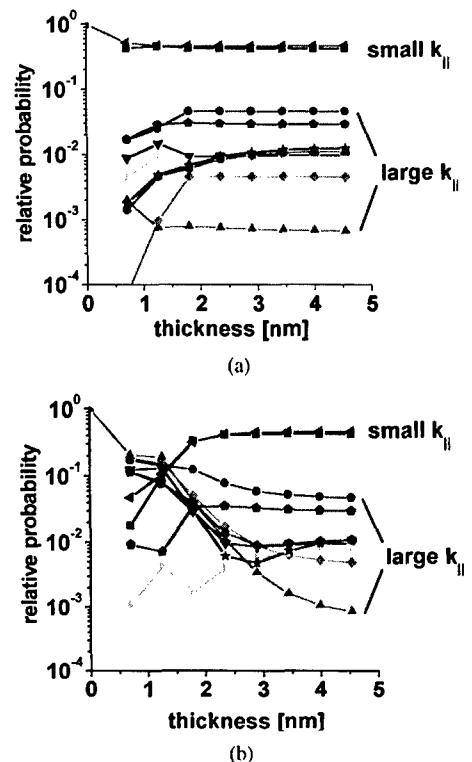


Figure 4. Relative probabilities that a Si bulk Bloch state with $E = 1.5$ eV, $\mathbf{k}_{\parallel}^{2 \times 2} \approx (0.2; 0)\text{nm}^{-1}$ and (a) small bulk \mathbf{k}_{\parallel} or (b) large bulk \mathbf{k}_{\parallel} leaves a 1.3 nm thin tridymite oxide in one of 10 possible states with the same $\mathbf{k}_{\parallel}^{2 \times 2}$ but different bulk \mathbf{k}_{\parallel} (at zero applied bias).

\mathbf{k}_{\parallel} . Figure 4(a) and (b) shows as a function of tridymite oxide thickness, the probabilities that a Bloch electron which hits the Si/ SiO_2 interface is scattered by the oxide into all 10 possible individual outgoing states. Three issues are important to note here: (i) after a tunneling distance of ≈ 5 nm, the probability distributions are identical, independent of the nature (small or large bulk \mathbf{k}_{\parallel}) of the incoming state, indicating that the oxide has lost all information on the incoming state, (ii) after a few nm tunneling distance, most of the tunneling electrons leave the oxide in states with small \mathbf{k}_{\parallel} : the oxide acts as a funnel in momentum space, (iii) only for t_{ox} greater than a certain critical value (about 3 nm for the tridymite oxide), the relative probabilities do not change anymore, i.e., all incident states see the same effective oxide barrier. Below this critical thickness, the barrier is thickness-dependent because of quantum mechanical interference effects. This is a possible explanation of the observed independence of oxide barrier heights on Si substrate orientation for oxides thicker than 5 nm (Weinberg 1982).

In the present formalism, \mathbf{k}_{\parallel} conservation is not only violated in transmission but also in reflection. In the present full-band framework, an incoming state i characterized by $(E, \mathbf{k}_{\parallel}^{n \times n}, k_{\perp}^i)$ is not exclusively scattered into a state with $(E, \mathbf{k}_{\parallel}^{n \times n}, -k_{\perp}^i)$ by specular reflection. Instead, reflection occurs in a set of p states $\{j\}$ with the same $(E, \mathbf{k}_{\parallel}^{n \times n})$ but various k_{\perp}^j . We introduce R_{ij} as the probability for such an individual scattering event $i \rightarrow j$. We have analyzed the R_{ij} for the interface between Si and the 2×2 tridymite oxide and various energies and parallel wavevectors, and tried to correlate them with the corresponding components of the group velocities parallel and normal to the interface. Our main findings are (i) the R_{ii} , relevant for specular reflection, vary strongly between typically 10^{-3} and 1 for the individual scattering events, (ii) averaged over a large set of scattering events, the R_{ii} and the other R_{ij} ($i \neq j$) are equally probable ($R_{ij} \approx 1/p$), which indicates that the scattering is (on the average) completely diffusive, and (iii) the results do not show an obvious dependence on energy. Since the present oxide models have perfectly flat interfaces, these effects are not related to any kind of interface roughness scattering but are rather pure band structure effects, caused by the mismatch of the two band structures on both sides of the interface, the anisotropy of the electronic structure of Si (compare Ham and Mattis (1960) and Price (1960)), and by the multiband nature of the Si band structure. This is important for transport along the Si-SiO₂ interface where interface scattering is a critical issue. We anticipate similar results for larger models that resemble the real amorphous oxide more closely. Also, parallel momentum breaking effects are expected to influence Fowler-Nordheim and thermal injection of electrons into SiO₂.

3.4. Direct Gate Currents in Sub-100 nm MOSFETs

As an application of the present scheme, we have calculated gate tunneling currents in prototypical MOSFET's (Antonianidis *et al.* 1999) with 50 and 90 nm gate lengths under realistic operating conditions by combining the tight-binding transmission coefficients with electron densities and data from full-band Monte Carlo transport simulations (Duncan, Ravaoli and Jakumeit 1998). For the present work, we have only considered distribution functions obtained from a transistor simulation in the saturation regime, with $V_{Gate-Source} = V_{Drain-Source} = 1.2$ V and 2.0 V for the 50 nm and 90 nm device, respectively. Tunneling

current densities $j(x)$ as a function of the coordinate x along the gate oxide have been obtained as

$$j(x) = -en(x) \sum_{\mathbf{k}, v_{\perp} > 0} \frac{v_{\perp}(\mathbf{k}) f(\mathbf{k}, x) T(\mathbf{k}, x)}{\sum_{\mathbf{k}'} f(\mathbf{k}', x)}. \quad (4)$$

Here, $n(x)$ is the electron density at the Si-SiO₂ interface, $f(\mathbf{k}, x)$ are the corresponding electron distribution functions, and $T(\mathbf{k}, x) \equiv T(E, \mathbf{k}_{\parallel}, x) \approx T_I(E, x)$. The sum in Eq. (4) includes only incident Si bulk states whose components of the group velocity normal to the interface, $v_{\perp}(\mathbf{k})$, are directed towards the oxide.

We find that for oxides with thicknesses smaller than ≈ 4 nm, gate leakage currents are dominated by tunneling of cold electrons in the source and drain contacts. As a consequence, the tunneling current densities (integrated over the entire gate length) decrease upon applying a drain-source voltage V_{DS} , i.e., increasing the energy of the electrons near drain (see Fig. 5). This can be understood by considering the factors that influence a gate tunneling current in a MOSFET: (i) the density $n(x)$, which is highest in the contacts and lowest in the channel, (ii) V_{DS} and V_{GS} , which influence the

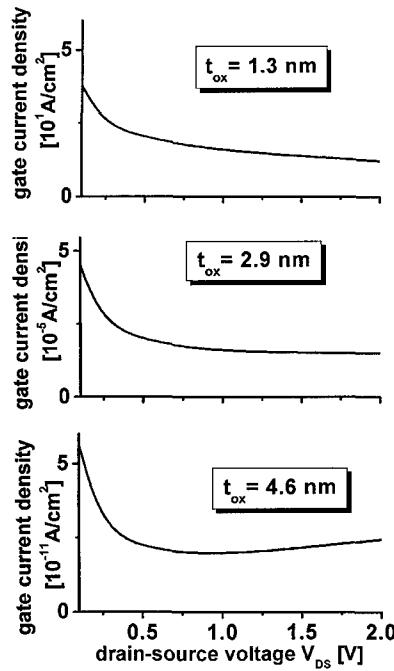


Figure 5. Drain-source voltage (V_{DS}) dependence of the average gate current densities in the 90 nm transistor considered in this paper. Data for various gate oxide thicknesses are shown; the gate-source voltage was chosen to be 2.0 V.

potential drop across the oxide, (iii) the shape of the distribution function $f(E, x)$, and (iv) the energy dependence of $T_I(E)$, which trivially favors tunneling of hot electrons (see Fig. 3). Reducing t_{ox} reduces the energy dependence of $T_I(E)$ exponentially, whereas the influence of factors (i)–(iii) is only moderately altered. However, for an extremely thin oxide, the energy dependence of $T_I(E)$ is only weak and counterbalanced mainly by factors (i) and (ii). Consequently, below a critical oxide thickness, the contribution to the tunneling current of the ‘many’ cold electrons in the contacts becomes larger than the contribution of the ‘few’ hot electrons. The critical value lies at ≈ 4 nm as can be seen from Fig. 5 for the 4.6 nm oxide: at V_{DS} greater than ≈ 1 V, the current increases again after its initial decrease, unlike in the cases with thinner oxides.

3.5. Influence of Elastic Defect-Assisted Tunneling on Gate Currents in Sub-100 nm-MOSFETS

It is interesting to see if the picture developed in the previous subsection still remains true when oxide defects are present. It is straightforward to create such defects (neutral oxygen vacancies, for example) in the present microscopic formalism by simply removing O atoms from the lattice. The Si atoms adjacent to the defect relax towards each other and form a bond whose antibonding level is believed to be energetically in a region relevant for electron tunneling (Pacchioni and Ieranò 1997, 1998, Blöchl and Stathis 1999). Since the exact position of this level is unknown and in a real oxide distributed over several tenths of an eV (Pacchioni and Ieranò 1997, 1998), we have treated it as a parameter (called E_{vac} hereafter, measured from the bottom of the conduction bands in the Si channel). We observe marked peaks in the oxide transmission $T_I(E)$ that can have a profound effect on gate currents, provided that the defects are located close to the center of the oxide and that $E_{vac} > \approx 0$. The full-width half-maximum values of the resonances range, for instance, for an 1.3 nm oxide from 4×10^{-4} eV to 5×10^{-2} eV as E_{vac} is varied from 0 to 2 eV. For a 2.9 nm oxide, we obtain respective values between 9×10^{-9} eV and 1×10^{-5} eV.

Subsequently, the elastic² gate leakage currents were recalculated including the O vacancies. The currents $j(x; n_{vac})$ at an arbitrary vacancy density n_{vac} (different from the reference density n_0 in the tight-binding calculations) were obtained using the interpolation formula $j(x; n_{vac}) = j(x; 0) + [j(x; n_0) - j(x; 0)]n_{vac}/n_0$. Interestingly, we find that for all possible combinations

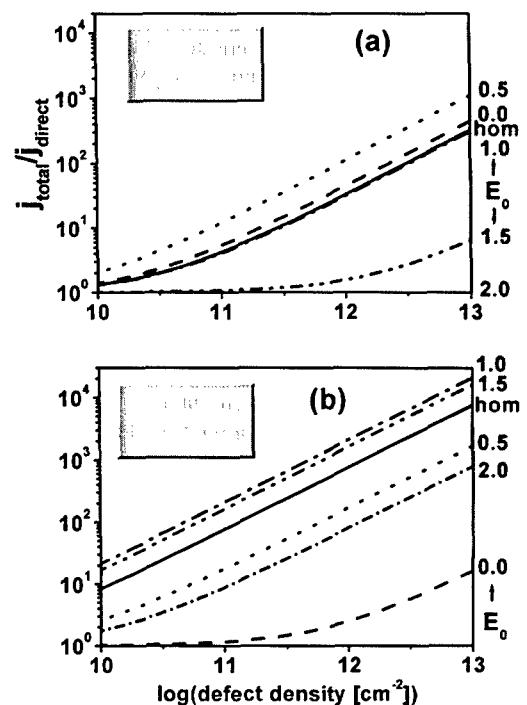


Figure 6. Dependence of the ratio of total (=direct and defect-assisted) and direct tunneling current density in the source regions of (a) a 50 nm and (b) a 90 nm MOSFET as function of the area density of the defects. Solid lines: homogeneous distribution of the defect energies in the interval $0 \text{ eV} < E < 2.5 \text{ eV}$; other lines: Gaussian distributions centered at various energies $E = E_0$ (in eV) with widths $\sigma = 0.3 \text{ eV}$. E is measured from the conduction band minimum in Si in all cases.

of E_{vac} and n_{vac} , the gate currents are still dominated by cold electrons originating in the contact regions. In Fig. 6(a) and (b), we show the ratio of total (direct and defect-assisted) and direct gate current densities from the source contact for the 50 nm transistor with a 1.3 nm oxide and the 90 nm transistor with a 2.9 nm oxide for homogeneous as well as a various Gaussian distributions of E_{vac} in energy space. It can be seen that the magnitude of the defect-induced current increase is very sensitive to the density and the energy distribution of the defects. For $n_{vac} > 10^{12} \text{ cm}^{-2}$, the enhancement can be as high as 2–3 orders of magnitude. Also, the resonant effects are somewhat less pronounced for the thinner oxide. This trend is in qualitative agreement with experimental results on virgin oxides (which are expected to be related to elastic tunneling channels (Ghetti *et al.* 2000) from Ghetti *et al.* (2000)). Given these observations, it appears to be possible that current enhancement effects of the same magnitude also occur for tunneling currents in one-dimensional MOS

structures and that averaging effects may hide possible sizeable elastic defect-assisted contributions.

We regard this work as the first steps toward the full understanding of oxide tunneling from a microscopic point of view.

Acknowledgments

Enlightening discussions with L.F. Register are acknowledged. This work has been supported by the ONR.

Notes

1. In the EM calculations, the effective masses derived from the tight-binding calculations have been consistently used.
2. We do not consider inelastic defect-assisted tunneling here since this would be beyond the scope of this paper. For very thin oxides, inelastic tunneling is expected to be of lesser importance since the dwell time of the electron on the defect is too short to induce the concomitant structural relaxation.

References

- Ando Y. and Itoh T. 1987. *J. Appl. Phys.* 61: 1497.
 Antonianidis O.A., Djomehri I.J., Jackson K.M., and Miller S. 1999. <http://www-mtl.mit.edu:80/Well/>.
 Av-Ron M., Shatzkes M., DiStefano T.H., and Gdula R.A. 1981. *J. Appl. Phys.* 52: 2897.
 Blöchl P.E. and Stathis J.H. 1999. *Phys. Rev. Lett.* 83: 372.
 Brar B., Wilk G.D., and Seabaugh A.C. 1996. *Appl. Phys. Lett.* 69: 2728.
- Dreizler R.M. and Gross E.K.U. 1990. *Density Functional Theory*. Springer, Berlin.
 Duke C.B. 1969. Tunneling in Solids. In: Seitz F., Turnbull D., and Ehrenreich H. (Eds.), *Solid State Physics*, Vol. 10. Academic Press, New York.
 Duncan A., Ravaioli U., and Jakumeit J. 1998. *IEEE Trans. Electron Devices* 45: 867.
 Franz W. 1956. In: Flügge S. (Ed.), *Handbuch der Physik*, Vol. 17. Springer, Berlin, p. 206.
 Fukuda M., Mizubayashi W., Kohno A., Miyazaki S., and Hirose M. 1998. *Jpn. J. Appl. Phys.* 37: L1534.
 Ghetti A., Alam M.A., Bude J., Sangiorgi E., Timp G., and Weber G. 2000. In: Proc. of the 4th Symposium on the Physics and Chemistry of SiO₂ and the Si-SiO₂ Interface, Toronto, Canada, p. 419.
 Grosso G. and Piermarocchi C. 1995. *Phys. Rev. B* 51: 16772.
 Ham F.S. and Mattis D.C. 1960. *IBM J.*, p. 143.
 International Technology Roadmap for Semiconductors (ITRS). 1999. Available at <http://public.itrs.net/>.
 Khairurrijal, Mizubayashi W., Miyazaki S., and Hirose M. 1999. Abst. of 1st Intern. Workshop on Dielectric Thin Films for Future ULSI Devices: Science and Technology, Tokyo, p. 11.
 Krieger G. and Swanson R.M. 1981. *Appl. Phys. Lett.* 39: 818.
 Maserjian J. 1974. *J. Vac. Sci. Technol.* 11: 996.
 Maserjian J. and Zaman N. 1982. *J. Vac. Sci. Technol.* 20: 743.
 Pacchioni G. and Ieranò G. 1997. *J. Non-Cryst. Solids* 21(6): 1.
 Pacchioni G. and Ieranò G. 1998. *Phys. Rev. B* 57: 818, and references therein.
 Price P.J. 1960. *IBM J.* p. 152.
 Städele M., Fischer B., Tuttle B., and Hess K. Solid State Electron., to appear.
 Städele M., Tuttle B., and Hess K. 2001. *J. Appl. Phys.* 89: 348.
 Strahberger C. 1999. Diploma Thesis, University of Regensburg, Germany.
 Strahberger C. and Vogl P. 1999. *Physica B* 272: 160.
 Weinberg Z.A. 1982. *J. Appl. Phys.* 53: 5052.
 Zhakarova A., Ryshii V., and Pesotzkii V. 1994. *Semicond. Sci. Technol.* 9: 41.



Full Quantum Simulation of Silicon-on-Insulator Single-Electron Devices

FREDERIK OLE HEINZ, ANDREAS SCHENK, ANDREAS SCHOLZE AND WOLFGANG FICHTNER

Integrated Systems Laboratory, ETH Zentrum, Gloriastrasse 35, CH-8092 Zürich, Switzerland

frederik.heinz@iis.ee.ethz.ch

Abstract. We present a method which extends the range of applicability of the domain decomposition approach to tunneling transport. Thereby we gain the ability to simulate e.g. structures with geometrically confined semiconductor quantum dots surrounded by very thin layers of dielectric or quantum dots that are defined through a combination of electrostatic forces and geometric confinement. Recently, experimental data of single electron devices on the 10 nm length-scale have become available, but due to the smallness of the devices detailed information on their geometry is hard to come by. Thus the simulations presented in this paper are intended as proof of principle rather than quantitative results for a real device. For predictive simulations more detailed knowledge of the experimental geometry is required.

Keywords: quantum dot, tunneling, domain decomposition, 3D, SOI, single electron transistor

1. Introduction

In the ongoing quest for ever smaller device dimensions and higher integration densities single electron devices might be able to play an important role. In this work we focus on silicon on insulator (SOI) single electron devices with direct tunneling as the dominant charge transport mechanism. The simulation geometry of an SOI single electron transistor (SET) is depicted in Fig. 1. It is derived from an experimental structure manufactured at the University of Tübingen (Augke *et al.* 2000). The diameter of the spherical quantum dot is 20 nm. The tunneling barriers reside in the constrictions in the silicon on either side of the central sphere.

2. Simulation Strategy

The quantum-mechanical charge density inside the device is computed by self-consistent solution of the Schrödinger–Poisson equations in effective mass approximation. In order to reduce the computational effort, the simulation volume is decomposed into domains of different dimensionality: source and drain contact regions are treated as two-dimensional electron gas; inside the quantum wires Schrödinger's equation is

adiabatically decomposed into a 1D array of 2D equations. Only inside the quantum dot the solution of the full 3D eigenvalue problem is necessary.

From the self-consistent single-particle wavefunctions in the diverse regions we then may obtain tunneling rates by Bardeen's transfer Hamiltonian method (cf. e.g. Payne (1986)). Subsequently we compute the linear response conductance of the device according to the approach by Beenakker (1991).

3. Adaptation of the Simulation Environment

The SIMNAD simulation environment (Scholze, Schenk and Fichtner 2000), developed at ETH, was originally designed for self-consistent conductance simulations of III–V single electron devices. In these devices quantum wires and dots were defined electrostatically by depletion of a 2DEG underneath metal electrodes. In contrast, SOI devices possess a fully three-dimensional geometry; electron localization is due to a combination of electrostatic forces and the geometrical confinement by the surrounding oxide. Also, in silicon we have to deal with a six-valley band structure with non-spherical iso-energy surfaces, whereas previously only spherical single-valley band structures

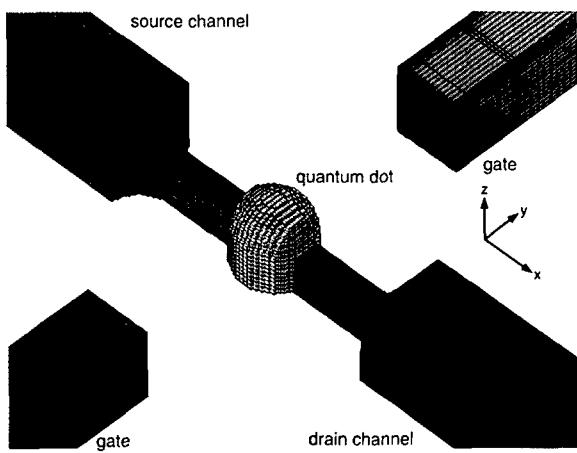


Figure 1. Simulation geometry of an SOI single electron transistor (oxide and substrate Si removed).

had to be considered. These differences necessitate several extensions to the simulation model.

3.1. Treatment of the Non-Spherical Six-Valley Band Structure of Silicon

In the effective mass approximation the six valleys of the silicon band-structure give rise to a Hamiltonian \mathcal{H}_{tot} operating on a Hilbert space of six component wave-functions. By neglecting inter-valley coupling, however, the full multi-valley Hamiltonian \mathcal{H}_{tot} may be decomposed into an (outer) direct sum of single-valley Hamiltonians

$$\mathcal{H}_{\text{tot}} \approx \bigoplus_{\alpha \in \{x, y, z\}} \mathcal{H}_{\alpha i}, \quad (1)$$

$$i \in \{+, -\}$$

$$\mathcal{H}_{\alpha i} = -\frac{\hbar^2}{2} \nabla \cdot \left(\left[\frac{1}{m_u^*} \right] \nabla \right) - eV_s[\rho], \quad (2)$$

where $\left[\frac{1}{m_u^*} \right]$ denotes the reciprocal effective mass tensor in a coordinate frame such that the main axis associated with its greatest mass component is along the α -axis, and $V_s[\rho]$ is the self-consistent potential brought about by ρ , the sum of the electron densities in all valleys. Thus the task of solving the 6-component Schrödinger equation is reduced to that of solving three scalar equations ($\mathcal{H}_{\alpha+} = \mathcal{H}_{\alpha-}$).

3.2. Handling of Moving Tunneling Barriers

In the SOI single electron transistor (SET) of Fig. 1 the definition of the quantum dot is due to a combination

of geometry and electrostatic effects. The variation of the transverse quantum kinetic energy along the transport direction is of the same order of magnitude as the depth of the electrostatic potential well inside the quantum dot region: depending on the gate voltage, a point may be found on either side of the tunneling barrier (cf. Fig. 2). Therefore, the simple strategy of defining a fixed Schrödinger box with Dirichlet boundary condition for the computation of the quantum dot levels breaks down: if the box is chosen too small, artificial boundary conditions will disturb the solution; if it is chosen too large, the Schrödinger solver will find solutions which are localized on the wrong side of the barrier ("spurious states").

This may be remedied by modifying the Hamiltonian for the 3D Schrödinger box in the spirit of

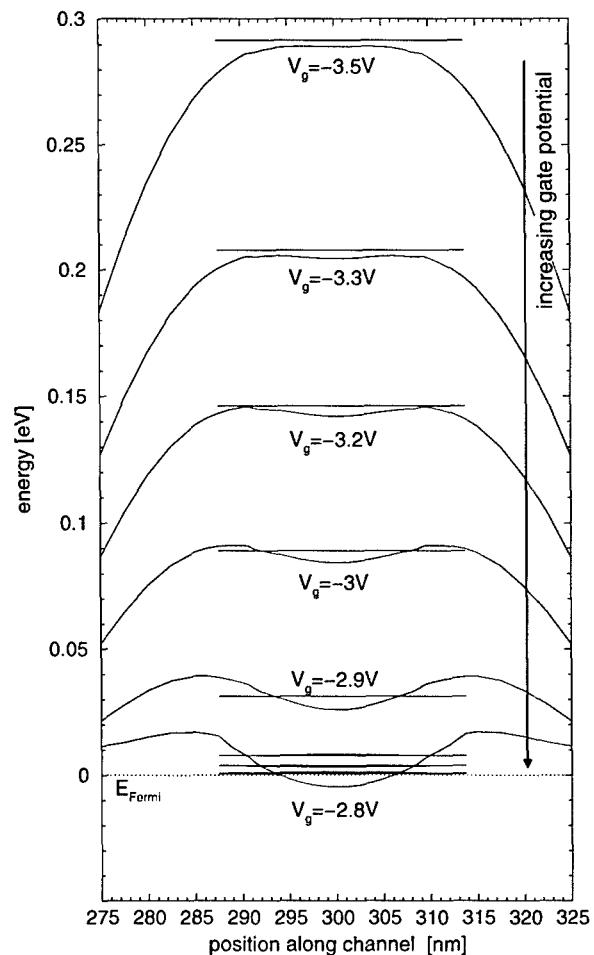


Figure 2. Quantum corrected conduction band energy and 3D eigen-energies at different gate voltages.

Bardeen's transfer Hamiltonian method: In 1D the transfer Hamiltonians $\mathcal{H}_{l/r}$ to either side of the barrier may be constructed by finding the position x_{\max} of the maximum of the barrier potential. The potential for $\mathcal{H}_{l/r}$ then is the unmodified potential $V(x)$ left/right of x_{\max} and $V(x_{\max})$ on the other side.

For higher dimensions this approach may be generalized by introducing the escape energy ϵ_{esc} : let x_0 be a point which is known to reside in the active dot volume (e.g. the minimum of the central potential well in the Schrödinger box). The escape energy then is the minimum energy at which there exists a classical trajectory from x_0 to infinity (i.e. to the boundary of the 3D Schrödinger domain, provided that it is chosen sufficiently large). Points x that may be reached classically from x_0 at energies smaller than ϵ_{esc} are said to be "inside" the active quantum dot volume Ω_{dot} . By leaving the potential inside Ω_{dot} unchanged and lifting it to at least ϵ_{esc} outside we may then construct a new Hamiltonian $\tilde{\mathcal{H}}_{\text{dot}}$ (cf. Fig. 3).

This construction is often successful in eliminating the spurious states. But in some situations it is too crude: it is blind to pure geometrical confinement. This shortcoming may be overcome by means of a quantum-corrected effective potential

$$\tilde{V}(x) := V(x) + \max_{|\hat{v}|=1} \epsilon_{\text{trans}}(x, \hat{v}), \quad (3)$$

where the transverse kinetic energy $\epsilon_{\text{trans}}(x, \hat{v})$ is defined as the expectation value of the kinetic energy operator for the lowest 2D state in a plane through x at normals to \hat{v} . The maximum is taken such that inside a constriction the dominant direction is selected. This new potential \tilde{V} then is used to construct a modified escape energy $\tilde{\epsilon}_{\text{max}}$ and active dot volume $\tilde{\Omega}_{\text{dot}}$ as above.

The improved quantum dot transfer Hamiltonian $\tilde{\mathcal{H}}_{\text{dot}}$ then is defined as

$$\begin{aligned} \tilde{\mathcal{H}}_{\text{dot}} := & -\frac{\hbar^2}{2} \nabla \left[\frac{1}{m^*} \right] \nabla \\ & + \begin{cases} V(x) : x \in \tilde{\Omega}_{\text{dot}} \text{ or } \tilde{V}(x) \geq \tilde{\epsilon}_{\text{esc}} \\ \tilde{\epsilon}_{\text{esc}} - \epsilon_{\text{trans}}(x) : \text{otherwise} \end{cases} \quad (4) \end{aligned}$$

The same method may also be used for quantum dots that are separated from neighboring semiconductor regions by a very thin layer of dielectric: here the Schrödinger box must extend some distance into the semiconductor on the other side of the dielectric such that the wave-function can recover from the Dirichlet

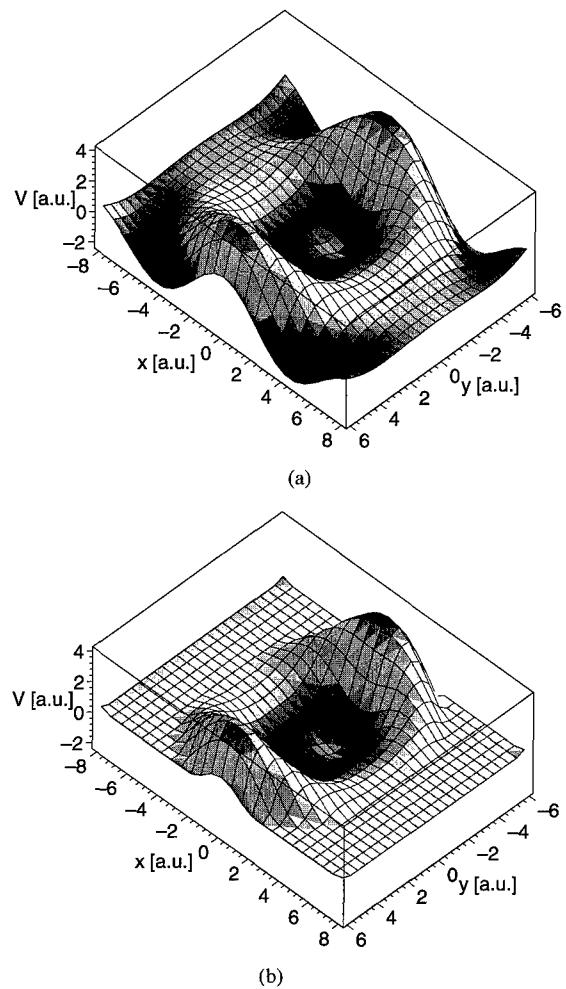


Figure 3. Example potential (a) unmodified (b) modified.

condition imposed on the box boundary; this will again bring about spurious states, that can be eliminated by the above method.

4. Results

With the modified transfer Hamiltonian $\tilde{\mathcal{H}}_{\text{dot}}$ the occurrence of spurious wave-functions could indeed be suppressed: all bound states are localized within the active dot volume, and there is almost no deformation due to the modified potential (cf. Fig. 4); the eigenenergies of the allowable single particle eigenstates were changed by less than 10 μeV (the numerical precision of the simulator).

The effective mass anisotropy has a pronounced effect on the shape of the wave-functions: depending

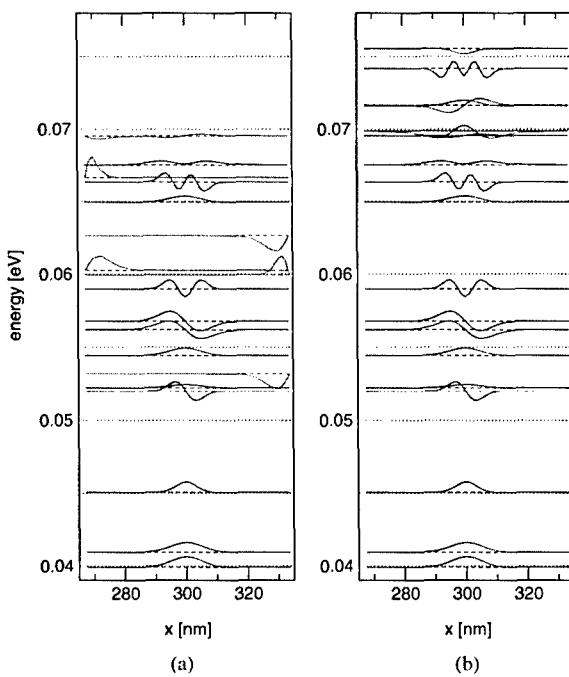


Figure 4. One-dimensional cuts through the eigenstates of (a) the original Hamiltonian \mathcal{H} ; (b) the improved transfer Hamiltonian $\tilde{\mathcal{H}}_{\text{dot}}$ [note the suppression of the spurious states by $\tilde{\mathcal{H}}_{\text{dot}}$].

on the orientation of the reciprocal effective mass tensor their spread along the transport direction varies so strongly that the tunneling rates of corresponding states in different valleys diverge by up to 8 orders of magnitude (cf. Fig. 5). The strong suppression of tunneling for $n_y = 2$ states relative e.g. to $n_z = 2$ states (where applicable the wave-functions are labelled by particle-in-a-box quantum numbers $n_x n_y n_z$) results from the symmetry of the structure in y -direction (the maximum of the channel wave-function coincides with a node of the dot wave-function) as opposed to the off-center position of the channel in z -direction: the quantum wire enters the quantum dot in the cylindrical bottom section, but is centered along the y -axis (cf. Fig. 1). The straight lines joining series of states (e.g. 111-211-311-411-511 for the $m_{\text{max}}^* = m_x^*$ orientation) correspond to an exponential increase of Γ with single particle energy.

The onset of conduction was found near a gate voltage of -2.5 V. Given that the simulation was modeled on a low resolution micrograph of the experimental structure together with the text description in Augke *et al.* (2000) this is in reasonable agreement with the experiment (experiment: first peak near -2.9 V). We find

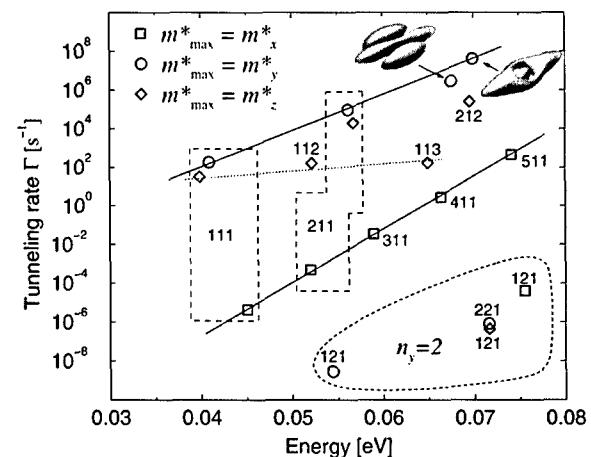


Figure 5. Source-dot tunneling rates of the single particle wavefunctions (particle-in-a-box quantum numbers $n_x n_y n_z$ shown where appropriate).

a spacing of the conductance peaks of about 100 mV, which also is not too far off from the experimental data.

5. Conclusions

In this paper we have been mostly concerned with technical difficulties that arise in the self-consistent quantum-mechanical simulation of SOI single electron devices. Now that they are overcome more detailed information on the device geometry is necessary in order to give true predictive power to our simulator. Only then will it be possible to decide the crucial question of whether a proposed device operates according to controllable conditions such as geometrical structure, or whether it depends on uncontrollable conditions such as an opportune configuration of individual dopant atoms, thus making reproducible production of such devices infeasible.

Acknowledgment

This work was funded by the European Union under contract number IST-1999-10828 (NANOTCAD).

References

- Augke R., Eberhardt C., Single F., Prins F.E., Wharam D.A., and Kern D.P. 2000. *Appl. Phys. Lett.* 76(15): 2065.
- Beenakker C.W.J. 1991. *Phys. Rev. B* 44: 1646.
- Payne M.C. 1986. The Institute of Physics, pp. 1145–1154.
- Scholze A., Schenk A., and Fichtner W. 2000. *IEEE Trans. Elec. Dev.* 47(10): 1811.



A 3-D Atomistic Study of Archetypal Double Gate MOSFET Structures

ANDREW R. BROWN, JEREMY R. WATLING AND ASEN ASENOV

*Device Modelling Group, Department of Electronics and Electrical Engineering, University of Glasgow,
Glasgow G12 8LT, Scotland, UK*

A.Brown@elec.gla.ac.uk

Abstract. The double gate MOSFET architecture has been proposed as a possible solution to allow the scaling of MOSFETs to the sub-30 nm regime, particularly due to its inherent resistance to short-channel effects. The use of lightly doped, or even undoped, channels means that such devices should be inherently resistant to random dopant induced fluctuations which will be one of the major obstacles to MOSFET scaling towards the end of the Si Roadmap. Random dopants within the channel are not, however, the only source of intrinsic fluctuations within MOSFETs at this scale. In this paper we investigate the impact of discrete dopants in the source and drain, individual charges within the active region and line edge roughness on the intrinsic parameter fluctuations in double gate MOSFETs.

Keywords: double gate MOSFET, atomistic, fluctuations, simulation, line edge roughness

1. Introduction

According to the updated 1999 edition of the International Roadmap for Semiconductors the MOSFET will reach 20 nm channel lengths by 2016. At the same time, theoretical studies indicate that the field effect action can be maintained to channel lengths below 10 nm where direct source-to-drain tunnelling may take over the gate control (Naveh and Likharev 2000). Properly scaled MOSFETs with 20 nm channel length and conventional architecture have already been demonstrated by leading semiconductor manufacturers (Chau 2001). It is, however, common wisdom that the scaling of the field effect transistor below this milestone requires intolerably thin gate oxide and unacceptably high channel doping, and therefore advocates a departure from the conventional MOSFET concepts. One of the most promising new device structures, scalable to dimensions below 10 nm, is the double gate MOSFET studied extensively in the last couple of years (Naveh and Likharev 2000, Chang *et al.* 2000, Ren *et al.* 2000). Theoretically the double-gate devices do not require channel doping to operate and therefore are considered to be inherently resistant to random dopant induced parameter fluctuations (Chang *et al.* 2000), which reach

an unacceptable level in their conventional counterparts (Asenov *et al.* 2001).

In this paper for the first time we carefully examine, using 3D atomistic simulations (Asenov *et al.* 1999), the resistance of sub 30 nm double-gate MOSFETs to intrinsic parameter fluctuations introduced by the discreteness of charge and atomicity of matter. We consider (i) the discrete dopants in the source and drain regions; (ii) individual charges in the active region of the device, associated with the background doping, fixed interface charge and trapped electrons; and (iii) the line edge roughness (LER) of the gate edge. Due to strong confinement effects in the thin silicon body of the double-gate MOSFETs the quantum mechanical (QM) effects dramatically affect the device electrostatics and are taken into account in our simulations using the well established density gradient (DG) formalism (Asenov *et al.* 2001, Rafferty *et al.* 1998).

We investigate double-gate MOSFETs as illustrated schematically in Fig. 1 with channel lengths ranging from 30 to 10 nm, and channel thickness between 5 and 1.5 nm (Ren *et al.* 2000). The importance of the quantum mechanical effects in such devices becomes apparent from Fig. 2. With the scaling of the channel length from 30 to 10 nm, and the corresponding

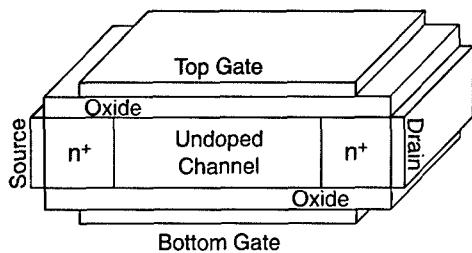


Figure 1. Schematic illustration of a double-gate MOSFET.

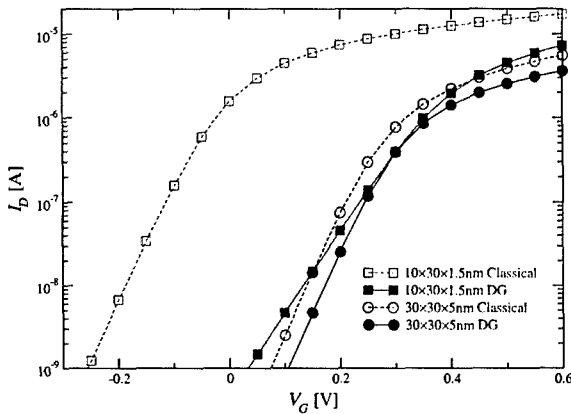


Figure 2. $I_D - V_G$ current characteristics for 10 and 30 nm channel length double-gate MOSFET, obtain from our classical and quantum simulations.

reduction in the silicon body thickness, the quantum mechanical threshold voltage shift increases from approximately 10 to 300 mV. Compared to the classical electron distribution, which demonstrates accumulation at both top and bottom interfaces, the quantum mechanical charge distribution peaks in the middle of the channel.

2. Intrinsic Parameter Fluctuations

2.1. Source/Drain Doping

While there may be no dopants within the channel region, Figs. 3 and 4 illustrate the impact of the unavoidable discrete dopants in the source/drain region on the potential and the electron distribution in a $30 \times 30 \times 5$ nm double-gate MOSFET. As shown in Table 1 the corresponding effective channel length fluctuations introduce threshold voltage fluctuations which increase from 0.66 mV to 1.07 mV as the device is scaled from 30 to 10 nm. These threshold voltage fluctuations are very small, as expected for the

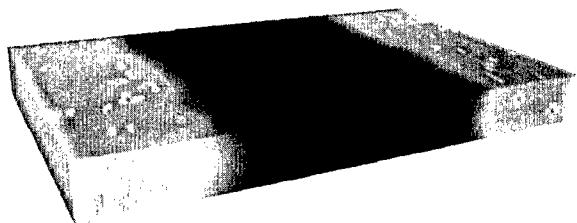


Figure 3. Electrostatic potential in a $30 \times 30 \times 5$ nm double-gate atomistic MOSFET at threshold.

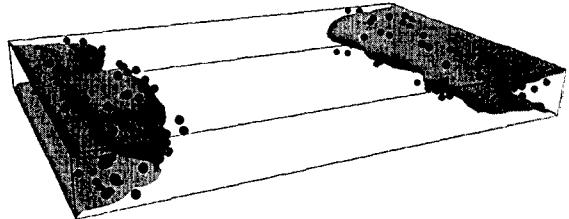


Figure 4. Electron equiconcentration surface in a $30 \times 30 \times 5$ nm double-gate atomistic MOSFET with location of dopants shown.

double-gate structure. The on-state current, however, does exhibit significant fluctuations, particularly for the shorter channel length device.

2.2. Fixed Charges in the Channel Region

Even with low doping levels for the channel region there is a possibility of at least one impurity (acceptor in the case of the n -channel MOSFET considered here) being present at a random location within the channel. Also, if an electron becomes trapped in an interface state or lattice defect it will also introduce a fixed negative charge to the channel region (Rals *et al.* 1984). These potential sources of additional charge will have an electrostatic effect on the channel potential, introducing a localised barrier to current flow, and also a shift in the threshold voltage (Asenov *et al.* 2000).

Table 1. Intrinsic parameter fluctuations in 10 nm and 30 nm double-gate MOSFETs due to atomistic doping in the source and drain. The threshold voltage is approximately 200 mV.

Channel dimensions ($L \times W \times T$)	Threshold voltage fluctuations (σV_T (mV))	Off-current fluctuations (σI_D (%))	On-current fluctuations (σI_D (%))
$10 \times 30 \times 1.5$ nm	1.07	9.56	7.13
$30 \times 30 \times 5$ nm	0.66	3.28	1.93

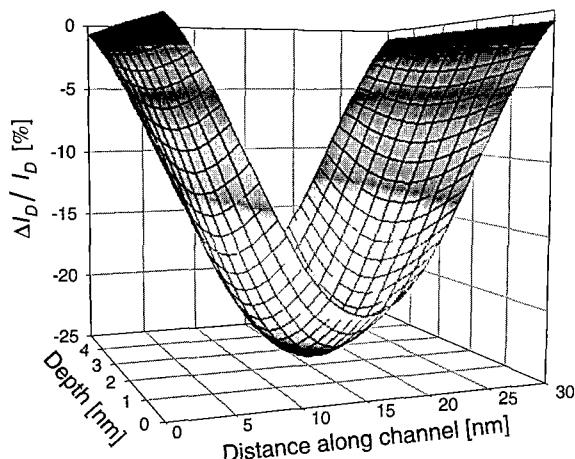


Figure 5. Percentage change in current when a charge is present at a particular location within the channel in a $30 \times 30 \times 5$ nm device.

The change in the drain current at threshold as a function of the position of a single negative charge in the device (either acceptor or trapped electron) is mapped for $30 \times 30 \times 5$ nm and $10 \times 10 \times 1.5$ nm double-gate MOSFETs in Figs. 5 and 6 respectively. This is for a vertical cross section running through the middle of the channel from source to drain. Due to the quantum distribution, resulting in the majority of current flowing in the plane through the middle of the device, a charge trapped in the centre of the channel produces the largest effect. The maximum reduction in the current increases from 24% in the 30×30 nm device to a staggering 65% in the 10×10 nm one, where the range of influence of the additional charge extends through much of the

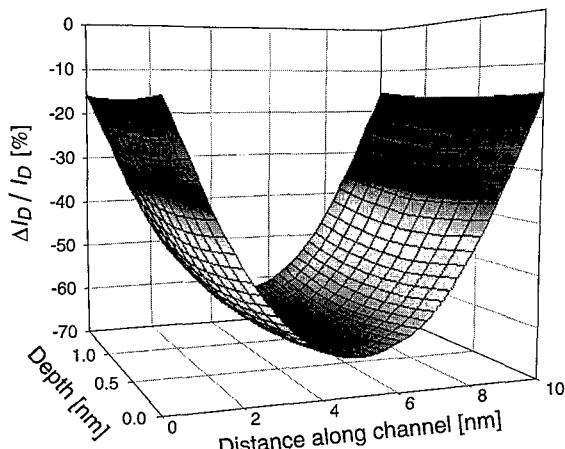


Figure 6. Percentage change in current when a charge is present at a particular location within the channel in a $10 \times 10 \times 1.5$ nm device.

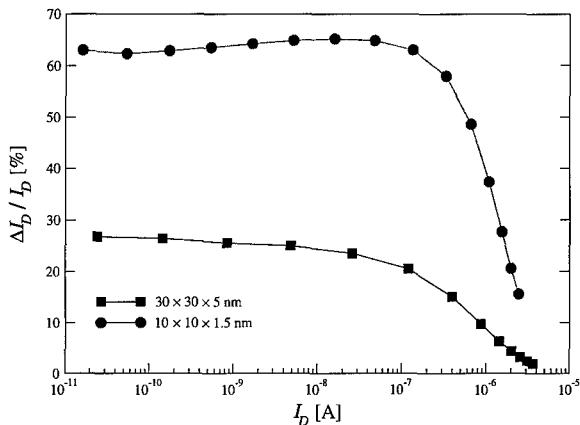


Figure 7. The reduction in current that is observed when a negative charge is present in the middle of the device, where the greatest influence occurs. The effect is most dramatic in the subthreshold region and reduces when above threshold.

device. This is further demonstrated by the fact that in the 10×10 nm device, where the thickness of the silicon is only 1.5 nm, the change in current has little dependence on the vertical position of the additional charge.

The relative reduction in the current as a function of the current itself is plotted in Fig. 7, for the ‘worst possible case’ scenario where the additional charge is in the centre of the channel. It is clear that the largest effect is in the subthreshold regime where the density of mobile charge in the channel is low. Above threshold, at higher drain currents in Fig. 7, the electrons in the channel screen the additional fixed charge, reducing its impact on the current flow, but not eliminating it completely.

2.3. Line Edge Roughness

Line edge roughness (LER), resulting from inherent tolerances in material specifications and in tool performance used in lithography processes is not a new phenomenon. Yet, the imperfections caused by LER effects have caused little worry for production lines over the years, since the critical dimensions of devices were more than an order of magnitude larger than the roughness. However, as the aggressive scaling of Si-MOSFETs continues to the sub-100 nm regime, LER does not diminish but constitutes a significant fraction of the gate length. At the dimensions being considered here for double-gate MOSFETs LER is of the same order of magnitude as the channel length, making it one of the prime problems for ULSI, where millions of

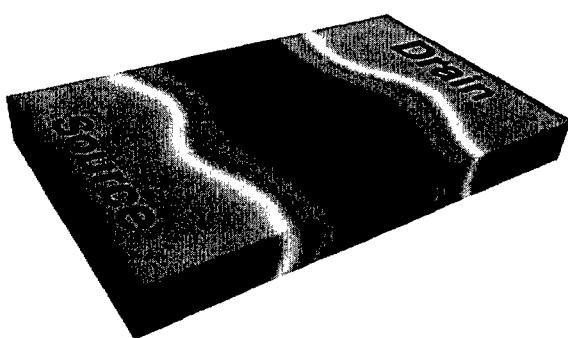


Figure 8. Potential distribution in a $30 \times 30 \times 5$ nm device illustrating the effect of line edge roughness (LER) on the source and drain junctions. $\Delta = 3$ nm and $\Lambda = 10$ nm.

devices are required to operate in very strict margins on a single chip.

Since the formation of LER is a stochastic event, a proper description and analysis of this phenomenon requires a strictly statistical approach. Realistic 'rough' lines produced by a lithography process can be statistically described by their RMS amplitude, Δ , and correlation length, Λ , which indicate the vertical and lateral extent of the roughness respectively.

An autocorrelation function for the random line is assumed, e.g. Gaussian or exponential. The power spectrum of this function, obtained by Fourier transform, is used for the amplitudes in a complex array. The phases are chosen randomly, which results in the random nature of the generated line which is obtained by inverse Fourier transform of the complex array (Kaya *et al.* 2001).

In our simulations we assume that the LER inherent in the fabrication process results in the p-n junctions in the MOSFET exhibiting the same rms amplitude and correlation length. We assume a Gaussian autocorrelation function for the line generation. The potential in a double gate MOSFET with LER is illustrated in Fig. 8.

Our analysis of published LER data from advanced lithography processes in various labs (Kaya *et al.* 2001) found that the value of LER (which is defined as being 3Δ) is 5 to 6 nm (i.e. $\Delta \approx 2$ nm) and, rather worryingly, is not reducing for shorter channel length technologies.

The standard deviation in threshold voltage, σV_T , as a result of the LER induced fluctuations are shown in Fig. 9, showing the dependence on rms amplitude, Δ . As one would expect, the fluctuations increase as Δ increases. It is clear that for the 10 nm device the fluctuations are massive, and with $\Delta = 3$ nm the standard deviation in threshold voltage is of the same order of magnitude as the threshold voltage itself. Figure 10

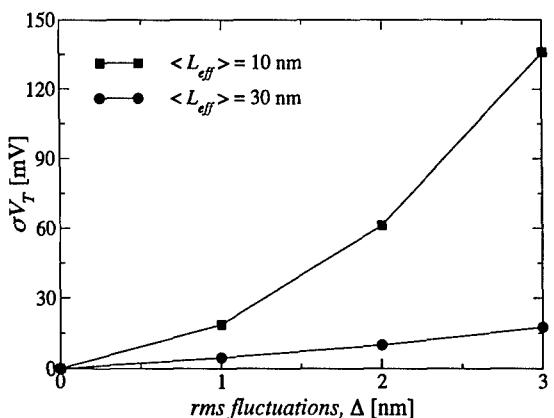


Figure 9. Standard deviation in threshold voltage, σV_T , due to fluctuations in line edge roughness of rms amplitude Δ , with $\Lambda = 20$ nm. Results for nominal channel lengths of 30 nm and 10 nm are shown. Device width is 30 nm in both cases.

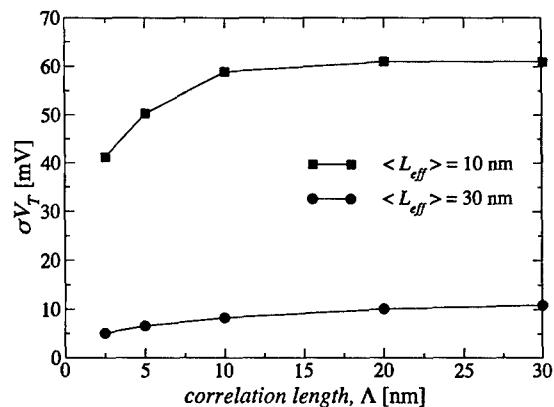


Figure 10. Standard deviation in threshold voltage, σV_T , due to fluctuations in line edge roughness of correlation length, Λ , with $\Delta = 2$ nm. Results for nominal channel lengths of 30 nm and 10 nm are shown. Device width is 30 nm in both cases.

shows the correlation length dependence of the fluctuations. The fluctuations increase with increasing correlation length, saturating when Λ is similar to the width of the device.

3. Conclusions

Double gate MOSFETs are a promising architecture for the scaling of devices to sub-20 nm dimensions. The undoped nature of the channel means that they are less susceptible to intrinsic parameter fluctuations due to the random number and location of dopants in the channel region, which will be a major problem with conventional MOSFET architectures.

However, we have shown that double gate MOSFETs are susceptible to other intrinsic sources of fluctuations. Random telegraph noise due to the trapping and de-trapping of electrons in lattice defects may result in large current fluctuations, which will be different for each device within an integrated circuit. The presence of even a single dopant within the channel will produce the same effect.

Line edge roughness inherent to current fabrication processes will be reflected in roughness of the p-n junctions of the device. As such, each device will have a different effective channel length and thus a different threshold voltage. If the present apparent limit of LER of approximately 5 nm is not reduced substantially then this will cause serious problems for devices with 10 nm channel lengths.

Acknowledgment

The authors would like to thank Savas Kaya for his work on random line generation. This work is

supported by SHEFC Research Development Grant VIDEOS and EPSRC grant GR/L53755.

References

- Asenov A., Balasubramaniam R., Brown A.R., Davies J.H., and Saini S. 2000. IEDM Tech. Digest, pp. 279–282.
- Asenov A., Brown A.R., Davies J.H., and Saini S. 1999. IEEE Trans. CAD Integ. Circuits and Systems 18: 1558.
- Asenov A., Slavcheva G., Brown A.R., Davies J.H., and Saini S. 2001. IEEE Trans. Electron Devices 48: 722–729.
- Chang L., Tang S., King T-J., Bokor J., and Hu C. 2000. IEDM Tech. Digest, pp. 719–722.
- Chau R. 2001. In: Proc. Si Nanoelectronics Workshop, pp. 2–3.
- Kaya S., Brown A.R., Asenov A., Magot D., and Linton T. 2001. In: Tsoukalas D. and Tsamis C. (Eds.), Simulation of Semiconductor Processes and Devices 2001. Springer-Verlag, Vienna, pp. 78–81.
- Naveh Y. and Likharev K.K. 2000. Superlattices and Microstructures 27(2/3).
- Rafferty C.S., Biegel B., Yu Z., Ancona M.G., Bude J., and Dutton R.W. 1998. In: Proc. SISPAD'98, pp. 137–140.
- Rals K.S., Scokpol W.L., Jakel L.D., Howard R.E., Fetter L.A., Epworth R.W., and Tennant D.M. 1984. Phys. Rev. Lett. 63: 228.
- Ren Z., Venugopal R., Data S., Lundstrom M., Jovanovic D., and Fossum J. IEDM Tech. Digest, pp. 715–718.



3-D Parallel Monte Carlo Simulation of Sub-0.1 Micron MOSFETs on a Cluster Based Supercomputer

ASIM KEPKEP AND UMBERTO RAVAIOLI*

Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ravaiol@uiuc.edu

Abstract. A full band, three-dimensional, Monte Carlo simulator for deep sub-micron Si MOSFET like devices has been developed, with the goal to obtain optimal performance on a parallel system built from a cluster of commodity computers. A short-range carrier-carrier and carrier-ion model has been implemented within this framework, using Particle-Particle Particle-Mesh (P3M) algorithm. Test simulations include the 90 nm “well-tempered MOSFET” for which measurements are available. Simulation benchmarks have identified several factors limiting the overall performance of the code and suggestions for improvements in these areas are made.

Keywords: Monte Carlo methods, MOSFET, parallelization

1. Introduction

The recent improvements in the performance of commodity computer hardware, along with the constant cost reduction, have created very favorable conditions for building cluster-based distributed parallel machines (Chien *et al.* 1999) reaching now performance levels of supercomputers designed just a few years ago. As a result, computer clusters have become very attractive for large-scale 3-D device simulation. Also, as the gate length of integrated MOSFETs is being scaled towards sub-0.1 μm dimensions, 3-D simulation is becoming necessary to understand the effects of geometry, and to properly model carrier-carrier interaction and granular doping effects (Mizuno, Okamura and Toriumi 1994, Wong and Taur 1993, Asenov 1998). There are areas of significant potential for improvement in functional density, by exploring 3-D geometry, where the functionality of the device can be enhanced by the availability of an additional degree of freedom (Lorenzini, Vissarion and Rudan 1999, Tanaka and Sawada 1996). Therefore, a 3-D Monte Carlo simulator capable of proper carrier-carrier and carrier-impurity interaction modeling would present a very valuable

capability to investigate performance of ultra-scaled Si-MOSFETs and other devices based on the MOS system.

Cluster based parallel computers have been demonstrated to achieve performance levels, once thought to be only achievable by using specially designed hardware and highly customized software. Monte Carlo methods in general lend themselves to parallel implementation due to relatively loose coupling between state variables associated with individual particles. The cluster environment has some unique characteristics, which had to be taken into account during the development of our 3-D simulator. Most important among these is the relatively high messaging latencies, stemming from the distributed nature of the computer. Task assignment to nodes must be made with sufficient care, so that the overhead due to communication among nodes does not jeopardize the overall performance of the simulator. The proper partitioning of tasks and the simulation domain decomposition over the cluster nodes has a significant impact on performance.

An important consideration during the development of the program has been to retain as much platform independence as possible, to ease future utilization of the codes in new computer architectures, without requiring a major redesign of the code. For this reason, standard

*To whom correspondence should be addressed.

FORTRAN and C languages have been used for coding. The communication and process initialization tasks are implemented using the MPI library (Gropp, Lusk and Skjellum 1994). This library offers a platform independent, communication interface. It is possible to approach optimal performance levels, with minimal or no changes to the simulator code, using an MPI library specifically written for a specific computer system. The simulator has been developed originally on a cluster of the National Center for Supercomputing Applications (NCSA), consisting of 96 double-processor INTEL Pentium III, working under the Windows NT operating system. During the later part of this work, another NCSA cluster has become available with the most advanced Intel processors and working under the Linux operating system. On both clusters, the processor nodes are connected with a high performance "Myrinet" network. This network has a peak bandwidth greater than 1 Gb/s and a low latency.

The main focus of this work has been the development of tools enabling simulation of small Si devices with the detailed physical models of full band Monte Carlo methods, in a high performance but affordable computational system, suitable for rapid turn-around simulation times in a modern fast-paced research and development environment. To achieve this goal, several different approaches to optimize parallelization have been investigated, including space domain decomposition, parallelization of Poisson's equation solver, at the level of matrix operations, and parallel implementation of bipolar simulation.

Reference test simulations for deep-scaled devices include the 90 nm and 25 nm gate length "well-tempered MOSFETs" (WTM) and a 25 nm MOSFET device with abrupt S-D doping profiles. The 90 nm structure has also been used to characterize the performance of the code since this structure is a better representative of the structures of interest in the immediate future. The MOSFET with abrupt doping profiles has been used for studying effects of grid size on the accuracy and performance of the program.

2. Parallel Monte Carlo

The 3-D device structure is decomposed into space sub-domains which are assigned to separate processors. In a typical MOSFET structure, it is efficient to "slice" the device with planes perpendicular to the interface, from source to drain, to minimize particle transfer from one domain to the other during the simulation (Kepkep and

Ravaioli 2001). As the Monte Carlo transport portion of the code is optimized, the solution of Poisson equation becomes increasingly the performance bottleneck. In our original implementation, Poisson equation was solved with a scalar conjugate gradients approach, on a single processor to which the charge information had to be sent. Due to the large number of Poisson solutions necessary, the time step being typically a fraction of femtosecond, we examined the possible option for parallelization of the solver. While a domain decomposition is also possible for Poisson equation, we decided to rather parallelize the actual matrix operation procedures required by the conjugate gradient iterations, resulting in a modest need for algorithm modification. Finally, holes are simulated as well in the device substrate, running a separate executable, so that electrons and holes are always simulated in a parallel fashion on separate processors.

Charge-charge interaction was implemented, following the approach developed for an earlier scalar version of the Monte Carlo simulator (Wordelman 2000). Both carrier-carrier and carrier-ion interaction can be evaluated, using the Particle-Particle-Particle-Mesh (P^3M) method (Hockney and Eastwood 1981). We performed a number of tests to assess the parallelization performance. The particle-particle interaction obviously introduces additional data exchanges among nodes, reducing the overall performance of the code. In order for all possible short-range pairs to be detected, all particles residing within the selected "short-range" action distance "a" of the domain boundaries must be exchanged. The distance "a" is chosen such that it spans at least two grid lines in any direction as in Wordelman (2000). If the slice thickness in the domain decomposition is set to four meshes, information on all the particles in the domain must be reported to the neighboring nodes on either side of the slice, as all particles are within the short-range interaction radius from the domain interfaces. As a result the runtime performance of the code suffered significantly, with 2 to 3 times increase in execution cost.

3. Simulation Results

Because of space limitations, we will only focus here on results for the largest of the WTM reference structures, with 90 nm channel length. This is also a good test to challenge the capabilities of the 3-D simulator, due to the size of the grid. The 2-D doping concentration was

imported from a file supplied as a part of the benchmark specifications, and it was kept constant in the third dimension, therefore creating a 3D sample of the device. This particular structure has a super steep retrograde channel doping as well as source and drain halo doping. The poly-Si gate is 300 nm thick and a gate oxide thickness of 4.5 nm. The active source and drain doping concentration is $7 \times 10^{19} \text{ cm}^{-3}$. A sketch of the device structure is shown in Fig. 1. The substrate doping is specified to be $3.9 \times 10^{15} \text{ cm}^{-3}$. The device has been simulated at $V_{gs} = V_{ds} = 2.0 \text{ V}$. This particular simulation was run on a cluster of 16 processors, with a discretization of $151 \times 140 \times 64$ grid points along x (channel direction), y (depth into substrate) and z (width), respectively. Slicing for domain decomposition is performed along the z -direction by assigning four grid points to each node.

A simulation time step of 0.1 fs was used, and a total of 120,000 time steps were simulated, with the first 50,000 time steps discarded as transient. The exact simulation time is not easy to estimate exactly in terms of CPU usage, but for this case the over wall clock time was about 8 hours. A range of results for carrier distribution, carrier velocity and potential profiles were compared with the results from the 2-D code and were found to be in agreement as expected, thus validating the proper implementation of the code. We just focus here on an aspect of the testing for the charge-charge interaction parallel implementation. The P³M technique was applied only to the evaluation of carrier-carrier interaction, since continuous doping profiles were specified. Figure 2 shows a comparison between the carriers distributions obtained for simulations with and without short-range interaction, near the channel/drain transition. The solution of Poisson equation is performed on a very fine non-uniform rectangular grid, in order to resolve sufficiently well the detail of carrier dynamics near the interface, where fields change rapidly and to faithfully describe the gradients of the specified

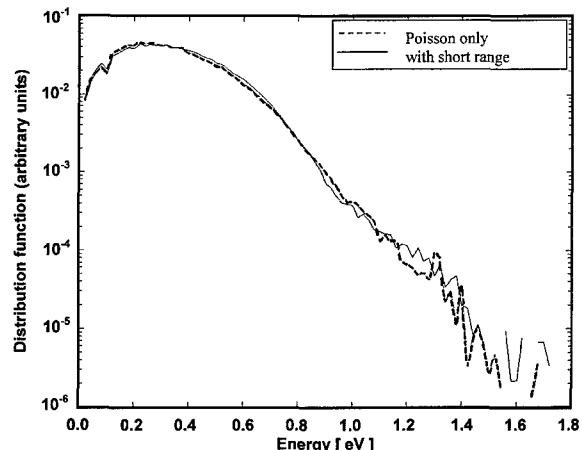


Figure 2. Carrier energy distribution example, obtained with and without detailed short-range charge-charge interaction for the 90 nm device.

doping profiles. Under these circumstances, we detect only a slight deformation of the energy carrier distribution when the P³M short-range interaction is added, resulting in some depression of the distribution that creates a very slightly enhanced tail at high energies. In 3-D Monte Carlo simulation, the solution of Poisson equation becomes a computational bottleneck, since it ends up requiring a sizable portion of the overall computational time. Depending on the simulation conditions, Poisson equation may end up accounting for more than 50% of the computational cost. Even at the high carrier concentrations achieved near the channel, only few carriers may reside in one given mesh and Poisson equation is able to capture nearly exactly the complete carrier Coulomb interaction. From such tests one can derive guidelines on when to apply the explicit but expensive evaluation of the short-range forces. For practical 3-D meshes, this may not be necessary, and the complete short-range evaluation should be reserved to treat cases where granular doping is specified. For ultra-small devices, smooth doping profiles may simply be unrealistic. With a distribution of actual ions, it is then possible to relax the restrictions on mesh spacing and use the coupling between Poisson equation and short-range force evaluation in an optimal way to reduce the overall cost of force evaluation. Inclusion of quantum correction would also lead to much softer carrier density profiles near the interface, further relaxing mesh requirements. We are now working on the 3-D implementation of quantum corrections based on a direct application of Schrödinger equation, demonstrated in 2-D codes.

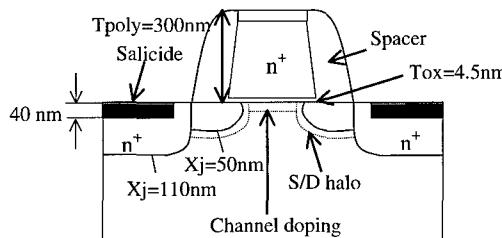


Figure 1. Schematic view of the 90 nm well-tempered MOSFET.

Acknowledgments

This work was supported by the Semiconductor Research Corporation, contract 99-NJ-726 and a DURINT Nanotechnology contract of the Army Research Office. The National Center for Supercomputing Applications (NCSA) provided access to the high performance cluster computers used for the simulations.

References

- Asenov A. 1998. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3D 'atomistic' simulation study. *IEEE Trans. Electron Dev.* 45: 2205–2513.
- Chien A., Lauria M., Pennington R., Showerman M., Iannello G., Buchanan M., Connelly K., Giannini L., Koenig G., Krishnamurthy S., Liu Q., Pakin S., and Sampemane G. 1999. Design and evaluation of an HPVM-based windows NT supercomputer. *Int. J. High Performance Computing Applications* 13: 201–219.
- Gropp W., Lusk E., and Skjellum A. 1994. *Using MPI: Portable Parallel Programming with Message Passing Interface*. MIT Press.
- Hockney R.W. and Eastwood J.W. 1981. *Computer Simulation Using Particles*. McGraw-Hill.
- Kepkep A. and Ravaioli U. 2001. Cluster-based parallel 3-D Monte Carlo device simulation. *VLSI Design* 13: 51–56.
- Lorenzini M., Vissarion R., and Rudan M. 1999. Three dimensional modeling of the erasing operation in submicron flash-EEPROM memory cell. *IEEE Trans. Electron Dev.* 46: 975–983.
- Mizuno T., Okamura J., and Toriumi A. 1994. Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's. *IEEE Trans. Electron Dev.* 41: 2216–2221.
- Tanaka J. and Sawada A. 1996. Simulation of a high performance MOSFET with quantum wire structure incorporating a periodically bent Si-SiO₂ interface. *IEEE Trans. Electron Dev.* 43: 2185–2198.
- Wong H.S. and Taur Y. 1993. Three dimensional 'atomistic' simulation of discrete random dopant distribution effects in sub-0.1 μm MOSFET's. In: Proc. IEDM'93, pp. 705–708.
- Wordelman C.J. 2000. Three-Dimensional Granular Monte Carlo Simulation of Semiconductor Devices. Ph.D. dissertation, University of Illinois at Urbana-Champaign.



Hole Transport in Orthorhombically Strained Silicon

F.M. BUFLER AND W. FICHTNER

Institut für Integrierte Systeme, ETH Zürich, Gloriastrasse 35, CH-8092 Zürich, Switzerland

Abstract. Linear and nonlinear transport of holes in orthorhombically strained Si to be used in vertical p-MOSFETs is theoretically analyzed. Strong mobility enhancements compared to unstrained Si by up to a factor of three is found at a Ge content of 40% in the SiGe pillar. The anisotropy in the three Cartesian directions is rather small and the saturation velocity remains unchanged. The enhanced material properties make orthorhombically strained Si attractive for device applications, although the improvements are not as strong as for biaxial tensile strain.

Keywords: orthorhombically strained Si, hole transport, Monte Carlo simulation

1. Introduction

The traditional way of achieving continued performance enhancement of silicon microelectronic devices has consisted for the last decades in downscaling the lateral dimensions of planar MOSFETs (metal-oxide-semiconductor field-effect transistors). However, this strategy is becoming increasingly difficult due to physical and technological problems such as the limited resolution of lithography. Therefore, also alternative approaches to improve device performance are currently being explored. Promising examples are vertical MOSFETs, where shorter channel lengths can be obtained by epitaxy, and planar heterostructure devices using biaxially strained Si with enhanced mobilities. In the case of p-MOSFETs, successful realizations of these two concepts can be found in Moers *et al.* (1999), Yang *et al.* (1999), Nayak *et al.* (1993), Rim *et al.* (1995) and Sugii, Yamaguchi and Nakagawa (2001), respectively. Recently, Liu *et al.* (1999) have reported the fabrication of a vertical n-MOSFET based on orthorhombically strained Si thus combining both methods, and improved electron drift velocities have been confirmed for this material by Monte Carlo simulation (Wang *et al.* 2000). It is the aim of this paper to address the corresponding situation in the complementary p-MOSFET structure with respect to the basic trans-

port properties, i.e. to compute the mobilities and drift velocities of holes in orthorhombically strained Si.

2. Model and Verification

The band structure is calculated by the non-local empirical pseudopotential method including spin-orbit interaction and the band energies are stored on an equidistant mesh in \mathbf{k} -space with a spacing of $1/96 \times 2\pi/a_i$ where the a_i denote the lattice constants in the three Cartesian directions. The scattering model comprises scattering of holes by optical phonons and by inelastic acoustic phonons (Bufler, Schenk and Fichtner 2001) allowing a numerical computation of the Ohmic drift mobility via the microscopic relaxation time. The resulting lattice-temperature dependence of the hole mobility in unstrained Si is compared in Fig. 1 with experimental mobility data (Green 1990, Ottaviani *et al.* 1975). Note, however, that the electric fields applied in the time-of-flight experiments (Ottaviani *et al.* 1975) were too large for the Ohmic regime at low lattice temperatures. Hence, these experiments lead to an underestimation of the Ohmic drift mobility as has already been reported previously (Bufler and Meinerzhagen 1998). We have therefore also performed Monte Carlo simulations at exactly the same field strengths as applied

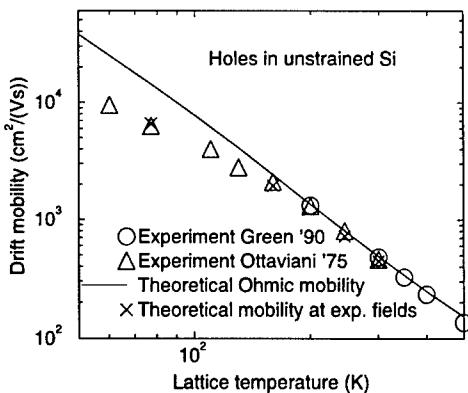


Figure 1. Lattice-temperature dependence of the theoretical hole drift mobility in unstrained Si compared to corresponding experimental data.

in the experiments. The results in Fig. 1 show good agreement between theory and experiments.

3. Orthorhombically Strained Si

As is illustrated in the schematic structure of a vertical MOSFET in Fig. 2, an orthorhombically strained Si layer can be obtained in the following way. First, a SiGe pillar is grown pseudomorphically on an unstrained Si substrate where the two in-plane lattice constants of SiGe, a_x and a_y , adopt the smaller bulk value of Si. This biaxial compressive strain leads in turn to a larger out-of-plane lattice constant a_z in the SiGe layer. In a second step, a Si layer is grown on the sidewall of the strained SiGe layer and on top of the unstrained

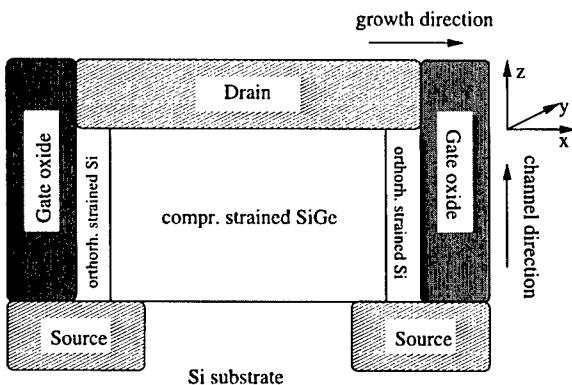


Figure 2. Schematic structure of the vertical MOSFET. The lattice constants a_x and a_y of the compressively strained SiGe pillar assume the value of the unstrained Si substrate which leads to an increase in a_z . a_y and a_z of the orthorhombically strained Si layer (=channel) coincide with the corresponding values in the Si substrate and the SiGe pillar, respectively, yielding a reduction of a_x .

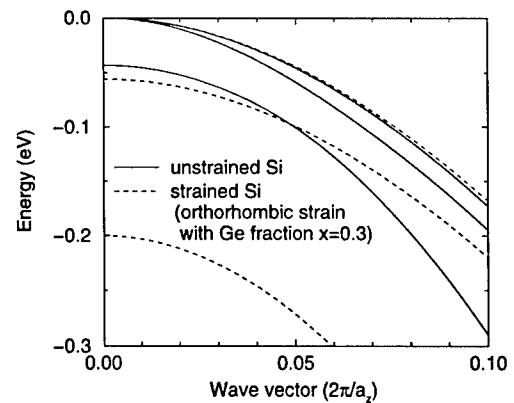


Figure 3. Valence-band energies along the k_z (channel) direction in unstrained Si and in orthorhombically strained Si (with a Ge fraction of $x = 0.3$ in the SiGe pillar).

Si substrate. As a consequence, the z -component of the lattice constant in the Si layer assumes the value of the corresponding component in the SiGe layer and the y -component equals the lattice constant of the Si substrate. This finally yields a smaller lattice constant in x -direction and the resulting orthorhombically strained Si layer forms the channel of the MOSFET. The strain-effect on the valence-band structure is displayed in Fig. 3 showing the three valence bands along the k_z direction, i.e. the channel direction. The degeneracy between the heavy-hole and the light-hole band at the Γ -point is lifted with the light-hole band being situated at the valence-band edge. The phonon scattering is as usually assumed to be unaffected by strain.

4. Results

The band splitting at the valence-band edge leads to a reduction of the density of states and therefore of the scattering rate. Hence, the strain enhances the mobility. This can be seen in Fig. 4 where the three diagonal components of the Ohmic drift mobility tensor are shown as a function of the Ge-content in the SiGe pillar. The anisotropy is relatively small, although the mobility in the growth direction is at higher Ge contents larger than the other components. As a general result, a strong mobility enhancement relative to unstrained Si is found, ranging from a factor of two at a Ge content of 20% up to a factor of about three at a Ge content of 40%. However, the increase in mobility above a Ge content of 20% is much weaker than in the case of biaxially strained Si where e.g. a value of roughly $2000 \text{ cm}^2/(\text{Vs})$ is reached at a Ge content of 30% (Bufler and

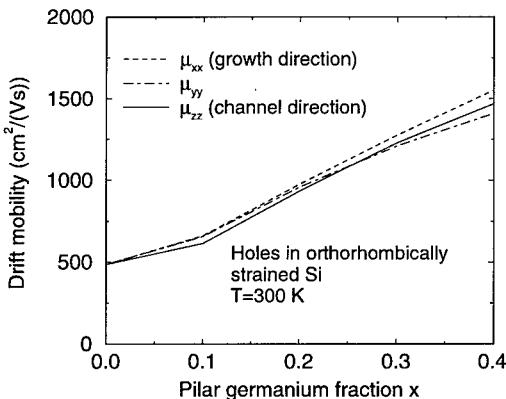


Figure 4. Diagonal components of the Ohmic drift mobility tensor for holes in orthorhombically strained Si at 300 K as a function of the Ge content in the SiGe pillar.

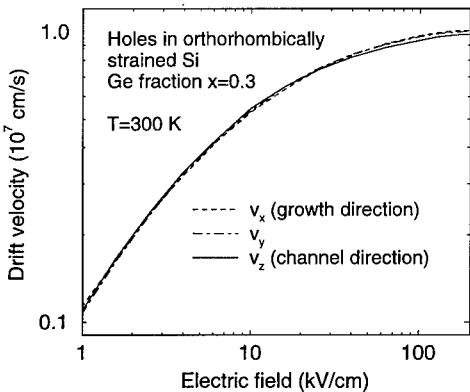


Figure 5. Velocity-field characteristics of holes in orthorhombically strained Si (with a Ge fraction of $x = 0.3$ in the SiGe pillar) at 300 K with the electric field oriented in x , y and z direction, respectively.

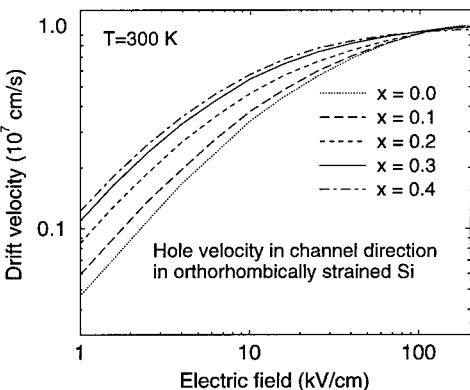


Figure 6. Velocity-field characteristics of holes in orthorhombically strained Si at 300 K with the electric field oriented in z (channel) direction for different Ge contents of the SiGe pillar.

Meinerzhagen 1998, Fischer and Hofmann 1999) as opposed to a value below $1500 \text{ cm}^2/(\text{Vs})$ in orthorhombically strained Si. Finally, the full-band Monte Carlo results for the velocity-field characteristics of holes in orthorhombically strained Si are presented in Figs. 5 and 6. In Fig. 5, the anisotropy of the drift velocity is illustrated for a Ge content of 30% and is found to be rather weak. Figure 6 reports the drift velocity in channel direction for several Ge contents of the SiGe pillar. The characteristics mainly reflect the tendency of the Ohmic drift mobility in Fig. 4, while the saturation drift velocities remain almost unchanged.

5. Conclusion

Linear and nonlinear transport of holes in orthorhombically strained Si has been theoretically investigated. Strong mobility improvements by factors between two and three relative to unstrained Si have been found for typical strain levels. From the point of view of basic transport properties this makes the material attractive for application in vertical p-MOSFETs, although the enhancement is not as strong as for biaxial strain.

Acknowledgment

We would like to thank M.M. Rieger for calculating the band structures.

References

- Bufler F.M. and Meinerzhagen B. 1998. J. Appl. Phys. 84: 5597.
- Bufler F.M., Schenk A., and Fichtner W. 2001. J. Appl. Phys. 90: 2626.
- Fischer B. and Hofmann K.R. 1999. Appl. Phys. Lett. 74: 2185.
- Green M.A. 1990. J. Appl. Phys. 67: 2944.
- Liu K.C., Wang X., Quinones E., Chen X., Chen X.D., Kencke D., Anantharam B., Chang R.D., Ray S.K., Oswal S.K., Tu C.Y., and Banerjee S.K. 1999. IEDM Tech. Dig. 63–66.
- Moers J., Klaes D., Tönnesmann A., Vescan L., Wickenhäuser S., Grabolla T., Marso M., Kordoš P., and Lüth H. 1999. Solid-State Electron. 43: 529.
- Nayak D.K., Woo J.C.S., Park J.S., Wang K.L., and MacWilliams K.P. 1993. Appl. Phys. Lett. 62: 2853.
- Ottaviani G., Reggiani L., Canali C., Nava F., and Alberigi-Quaranta A. 1975. Phys. Rev. B 12: 3318.
- Rim K., Welser J., Hoyt J.L., and Gibbons J.F. 1995. IEDM Tech. Dig. 517–520.
- Sugii N., Yamaguchi S., and Nakagawa K. 2001. Semicond. Sci. Technol. 16: 155.
- Yang M., Chang C.-L., Carroll M., and Sturm J.C. 1999. IEEE Electron Device Lett. 20: 301.
- Wang X., Kencke D.L., Liu K.C., Tasch A.F. Jr., Register L.F., and Banerjee S.K. 2000. J. Appl. Phys. 88: 4717.



Empirical Pseudopotential Method for the Band Structure Calculation of Strained-Silicon Germanium Materials

SALVADOR GONZALEZ AND DRAGICA VASILESKA

*Department of Electrical Engineering and Center for Solid State Electronics Research,
Arizona State University, Tempe, AZ 85287-5706, USA*

ALEXANDER A. DEMKOV

Physical Sciences Research Labs, Motorola, Inc., 7700 S. River Parkway, Tempe, AZ 85284, USA

Abstract. The band structure of strained-silicon germanium ($\text{Si}_{1-x}\text{Ge}_x$) is calculated as a preliminary step in developing a full band Monte Carlo (FBMC) simulator. The band structure for the alloy is calculated using the empirical pseudopotential method (EPM) within the virtual crystal approximation (VCA). Spin-orbit interaction is included into the calculation via the Löwdin quasi-degenerate perturbation theory, which significantly reduces the computation time. Furthermore, strain is included by utilizing basic elastic theory. Ultimately, the band structure for strained $\text{Si}_{1-x}\text{Ge}_x$ is calculated at various germanium concentrations.

Keywords: strained-silicon germanium, band structure calculation, empirical pseudopotential method, virtual crystal approximation

1. Introduction

Current market forces demand that the semiconductor industry produce faster integrated circuits (ICs) with high functionality at a low cost. One way of achieving this trend is to scale the device geometry. The industry is quickly reaching the physical limitations of small devices, however. In metal-oxide semiconductor (MOS) transistors, for example, thin oxides give way to high gate leakage currents. Increased short-channel effects (SCE) also impede performance improvements. One solution that replaces device scaling is the introduction of new materials.

For this purpose, strained-silicon or strained-silicon germanium ($\text{Si}_{1-x}\text{Ge}_x$) material systems have received much attention as possible candidates for improving performance of existing Si technology (Iyer *et al.* 1989, Harame *et al.* 1995a, b, Cressler 1995). This trend has been made possible via recent innovations in molecular beam epitaxy (MBE) growth techniques that allow for relatively easy growth of $\text{Si}_{1-x}\text{Ge}_x$ on $\text{Si}_{1-y}\text{Ge}_y$ substrates. Furthermore, $\text{Si}_{1-x}\text{Ge}_x$ can be integrated into

existing Si technology without the need for significant factory retooling.

For strained- $\text{Si}_{1-x}\text{Ge}_x$ material systems, the full band structure is required in order to capture the band splitting and warping, especially near the valence band maximum at the zone center (Γ). To this end, the full band structure of strained-silicon germanium is calculated using the EPM with spin-orbit interaction included.

2. Empirical Pseudopotential Method

The pseudopotential method is based on the Phillips-Kleinman cancellation theorem (Phillips and Kleinman 1959), which provides justification why the electronic structure can be described using a nearly-free electron model and weak potentials. For this purpose, the pseudopotential Hamiltonian can be written as

$$H = -(\hbar^2/2m)\nabla^2 + V_P(r), \quad (1)$$

where $V_P(r)$ is the smoothly-varying pseudopotential (Cohen and Bergstresser 1966). Because the crystal

potential is periodic, the pseudopotential is also a periodic function and can be expanded into a Fourier series over the reciprocal lattice to obtain

$$V_P(\mathbf{r}) = \sum_{\mathbf{G}} S(\mathbf{G}) V_{ff}(\mathbf{G}) e^{i\mathbf{G} \cdot \mathbf{r}}, \quad (2)$$

where $S(\mathbf{G})$ is the structure factor and $V_{ff}(\mathbf{G})$ is the pseudopotential form factor, which is defined as twice the inverse Fourier transform of the atom potential. For diamond-lattice materials, the structure factor is defined as $S(\mathbf{G}) = \cos(\mathbf{G} \cdot \boldsymbol{\tau})$, where $\boldsymbol{\tau} = a(1/8, 1/8, 1/8)$ is the atomic basis vector defined in terms of the lattice constant a when the coordinate origin is taken to be halfway between the basis atoms.

Because the pseudopotential in a crystal lattice is periodic, it follows that the pseudo-wave function corresponding to (1) is also periodic and can be expressed as a Bloch function, which consists of a plane-wave part and a cell periodic part. The cell periodic part, in turn, can be expanded into a Fourier series over the reciprocal lattice. By substituting the expanded pseudo-wave function and the pseudopotential defined by (2) into the Schrödinger wave equation, the Hamiltonian matrix results and is defined as

$$H_{i,j} = \begin{cases} \frac{\hbar^2}{2m} |\mathbf{k} + \mathbf{G}_i|^2, & i = j \\ V_{ff}(|\mathbf{G}_i - \mathbf{G}_j|) \cos[(\mathbf{G}_i - \mathbf{G}_j) \cdot \boldsymbol{\tau}], & i \neq j \end{cases} \quad (3)$$

where \mathbf{G} is a reciprocal lattice vector and \mathbf{k} is a wave vector lying within the first Brillouin zone. The solution to the energy eigenvalues and corresponding eigenvectors can then be found by diagonalizing the Hamiltonian matrix. For this work, 137 plane waves, each corresponding to reciprocal lattice vectors up to and including the 10th-nearest neighbor from the origin, were used to expand the pseudopotential. The publicly available eigenvalue solver LAPACK was used to diagonalize the Hamiltonian matrix.

3. Spin-Orbit Interaction

To develop a more refined picture of the energy bands, the spin-orbit interaction must be included into the pseudopotential calculation. In the context of electronic structure theory, the spin-orbit interaction serves to split degenerate energy levels. This influence is most pronounced for the valence band maxima near the Brillouin zone center.

For an electron orbiting a nucleus, which produces a spherically symmetric potential V , the spin-orbit interaction is calculated using Einstein's special theory of relativity to obtain

$$H_{SO} = \frac{\hbar}{4m^2c^2} \left[\frac{1}{r} \frac{\partial V}{\partial r} \right] \mathbf{L} \cdot \boldsymbol{\sigma}, \quad (4)$$

where \hbar is the reduced Planck constant, m is the electron's rest mass, c is the speed of light, \mathbf{L} is the electron's orbital angular momentum and $\boldsymbol{\sigma}$ is the Pauli spin tensor.

It may be tempting to add the H_{so} term from (4) directly to (1) and obtain the solution by diagonalizing the total Hamiltonian. This would not be the correct way to proceed, however, given that the pseudo-wave function corresponding to (1) is a spinless quantity. When spin is included into the problem, the crystal wave function becomes a (2×1) -spinor. By using shorthand subscript notation for spin, the spin Hamiltonian is given by

$$H_{m'm'\mathbf{k}'\sigma';m\mathbf{k}\sigma} = E_{m\mathbf{k}}^0 \delta_{m'm';m\mathbf{k}\sigma} + \langle \Psi_{m'm'\mathbf{k}'\sigma'} | H_{SO} | \Psi_{m\mathbf{k}\sigma} \rangle, \quad (5)$$

where m' is the row index and m is the column index and $\sigma = \pm 1$ is the Pauli spin index corresponding either to the spin up or spin down state. In this way, the spin Hamiltonian can be constructed using the spinless eigenvalues as the diagonal elements and including the spin-orbit interaction as a perturbation. It has been shown (Saravia and Brust 1968), however, that for states containing l -symmetry already included in the core states ($2p$ core states for Si and $3p$ core states for Ge), the perturbation in (5) can be written as a double summation over the reciprocal lattice vectors

$$\begin{aligned} & \langle \Psi_{m'm'\mathbf{k}'\sigma'} | H_{SO} | \Psi_{m\mathbf{k}\sigma} \rangle \\ &= -i\lambda_p \sum_{\mathbf{G}_i, \mathbf{G}_j} a_{m'\mathbf{k}'}^*(\mathbf{G}_i) a_{m\mathbf{k}}(\mathbf{G}_j) S(\mathbf{G}_j - \mathbf{G}_i) \\ & \quad \times F_p(\mathbf{k} + \mathbf{G}_i) F_p(\mathbf{k} + \mathbf{G}_j) \\ & \quad \times [\mathbf{e}(\mathbf{k} + \mathbf{G}_i) \mathbf{e}(\mathbf{k} + \mathbf{G}_j)] \cdot \boldsymbol{\sigma}_{\sigma'\sigma} \end{aligned} \quad (6)$$

where $S(\mathbf{G})$ is the structure factor, λ_p is a free parameter used to adjust the energy splitting, F_p is a function associated with p -core states, $\mathbf{e}(\mathbf{k})$ is a unit vector in the \mathbf{k} direction and σ is related to the Pauli spin matrices.

Including the spin-orbit interaction serves to double the size of the Hamiltonian matrix. In addition, each spin-orbit matrix element is calculated as the double summation over the reciprocal lattice vectors \mathbf{G}_i and \mathbf{G}_j , as seen in (6). As a result, the Hamiltonian is

computationally expensive to calculate, especially since there are 137 reciprocal lattice vectors employed in the EPM. To minimize the computational cost, Löwdin's quasi-degenerate perturbation theory is applied.

Löwdin's perturbation technique serves to reduce the size of the eigenvalue problem by "concentrating" the information in the initial Hamiltonian matrix to obtain a smaller matrix (Löwdin 1951). Löwdin uses the variational principle to arrive at a perturbation formula, which gives the influence of the higher-lying (class B) states on the lower-lying (class A) states. The class B states are eliminated through a process of iteration to obtain

$$\sum_n^A (U_{mn} - E \delta_{mn}) c_n = 0, \quad (7)$$

where

$$U_{mn} = H_{mn} + \sum_{\alpha}^B \frac{H'_{m\alpha} H'_{\alpha n}}{E - H_{\alpha\alpha}} + \sum_{\alpha}^B \sum_{\beta}^B \frac{H'_{m\alpha} H'_{\alpha\beta} H'_{\beta n}}{(E - H_{\alpha\alpha})(E - H_{\beta\beta})} + \dots \quad (8)$$

The first term H_{mn} in (8) is a matrix element, which corresponds to an A-class state, in the initial matrix. The subsequent terms correspond to the influence of the B-class states, which are treated as a perturbation here, on the A-class states. For this work, the first two terms in (8) are included in the calculation of U_{mn} .

The benefits of using the quasi-degenerate perturbation theory are: (a) one does not need to solve an eigenvalue problem of size $2N$ when spin is included into the problem and (b) degenerate and non-degenerate states are treated on an equal footing, which means that there is no need to first lift the degeneracy before applying the perturbative approach. Within this scheme, the degeneracy of the states is lifted via the introduction of the effective matrix element U_{mn} .

Since (8) is calculated through a process of iteration, the value E is introduced into the expression. For this work, E is estimated to be the average energy of the class-A states. Furthermore, 60 class-A states are used to achieve an eigenvalue convergence within 5 meV for states near the valence band maxima. For the case that the problem is solved exactly, each \mathbf{k} -point requires approximately 10.5 sec of central processing unit (CPU) time on a 500 MHz Pentium III microprocessor. Using the Löwdin perturbation technique with 60 class-A states only 1 sec of CPU time is required to solve for the energy spectrum at each \mathbf{k} -point.

Finally, the spin-orbit parameter λ_p in (6) that produces the appropriate spin-orbit splitting, i.e. 44 meV for Si and 300 meV for Ge, is determined by linear interpolation. The value for Si is $\lambda_p^{Si} = 0.00156 \text{ eV}\cdot\text{cm}^3$, and the value for Ge is $\lambda_p^{Ge} = 0.0112 \text{ eV}\cdot\text{cm}^3$.

4. Silicon Germanium Alloy

The elemental semiconductors silicon (Si) and germanium (Ge) are isoelectronic. As a result, their chemical and electronic properties are similar. Si and Ge are the only group-IV elements that are completely miscible. It is thus possible to form a solid solution of one element in the other to obtain a silicon germanium ($\text{Si}_{1-x}\text{Ge}_x$) alloy. The material properties vary gradually over the entire range. The lattice constant, for example, varies nearly linearly over the range of x . This fact is quantified by Vegard's rule (Vegard 1921) which states that the bulk lattice constant is given by

$$a^{\text{Si}_{1-x}\text{Ge}_x}(x) = a^{\text{Si}}(1 - x) + a^{\text{Ge}}x. \quad (9)$$

Like its constituent elements, the bulk silicon germanium alloy crystallizes in a diamond lattice, which is characterized by face-centered cubic (FCC) symmetry. From the definition of alloy, Ge atoms substitute for Si atoms randomly throughout the crystal, in proportion to the Ge concentration, x .

Because the material properties vary gradually over the range of Ge concentrations, it is possible to apply the virtual crystal approximation (VCA) to include alloy information. A silicon germanium alloy can be approximated as a FCC lattice of "hybrid" atoms. It then follows from the VCA that all the alloy parameters in the EPM can be interpolated with respect to the Ge concentration, x .

5. Strained-Silicon Germanium Alloy

To obtain a strained-silicon germanium alloy, $\text{Si}_{1-y}\text{Ge}_y$ is pseudomorphically grown on top of a $\text{Si}_{1-y}\text{Ge}_y$ substrate. The in-plane lattice constant of the growth layer conforms to the substrate, making the in-plane lattice constant different than its bulk value. From elastic theory it follows that the growth layer experiences biaxial strain in the direction of the growth plane. The in-plane strain condition can be expressed as

$$\varepsilon_{//} = \frac{a^{\text{Si}_{1-y}\text{Ge}_y} - a_{//}^{\text{Si}_{1-x}\text{Ge}_x}}{a_{//}^{\text{Si}_{1-x}\text{Ge}_x}}, \quad (10)$$

which is a relative change in lattice constants due to the stress.

Elastic theory predicts that the growth layer will respond in the direction normal to the growth interface plane in order to minimize its elastic energy (Rieger and Vogl 1993). To satisfy minimum energy, the transverse strain ε_{\perp} is given by

$$\varepsilon_{\perp} = -2 \frac{c_{12}}{c_{11}} \varepsilon_{\parallel}, \quad (11)$$

where c_{12} and c_{11} are the elastic constants of $\text{Si}_{1-x}\text{Ge}_x$, also calculated within the VCA. It then follows that the strain of the $\text{Si}_{1-x}\text{Ge}_x$ growth layer is given by the following second-rank tensor

$$\overset{\leftrightarrow}{\varepsilon} = \begin{bmatrix} \varepsilon_{\parallel} & 0 & 0 \\ 0 & \varepsilon_{\parallel} & 0 \\ 0 & 0 & \varepsilon_{\perp} \end{bmatrix}, \quad (12)$$

for which the elements are defined by (10) and (11). The vanishing off-diagonal elements in (12) indicate that there is no shear strain in the system. The system undergoes a deformation along the principle axes only.

The key elements used in the implementation of the EPM are the reciprocal lattice vectors. To include strain into the EPM, it is necessary to apply strain to the reciprocal lattice vectors. To do this, strain is first applied to the primitive lattice vectors of the direct space to obtain the strained lattice vector

$$\overset{\leftrightarrow}{\mathbf{a}}'_{\alpha} = (\overset{\leftrightarrow}{1} + \overset{\leftrightarrow}{\varepsilon}) \mathbf{a}_{\alpha}, \quad (13)$$

where $\overset{\leftrightarrow}{1}$ is the unit tensor and \mathbf{a}_{α} is an unstrained lattice vector. The atomic basis vector, τ , is also transformed under (13). Once the strained direct lattice vectors are calculated, the strained reciprocal lattice vectors \mathbf{G}'_{α} are calculated as

$$\mathbf{G}'_{\alpha} = 2\pi \frac{\mathbf{a}'_{\beta} \times \mathbf{a}'_{\gamma}}{\mathbf{a}'_1 \cdot (\mathbf{a}'_2 \times \mathbf{a}'_3)}, \quad (14)$$

The pseudopotential is then expanded over the strained reciprocal lattice vectors to include strain into the EPM.

6. Results

The band structure is calculated for strained- $\text{Si}_{1-x}\text{Ge}_x$ ($x = 40\%$) on a Si substrate (Fig. 1). The pseudopotential form factors for Si and Ge are taken from Chelikowsky and Cohen (1974) and Saravia and Brust (1968), respectively. A key feature in the band

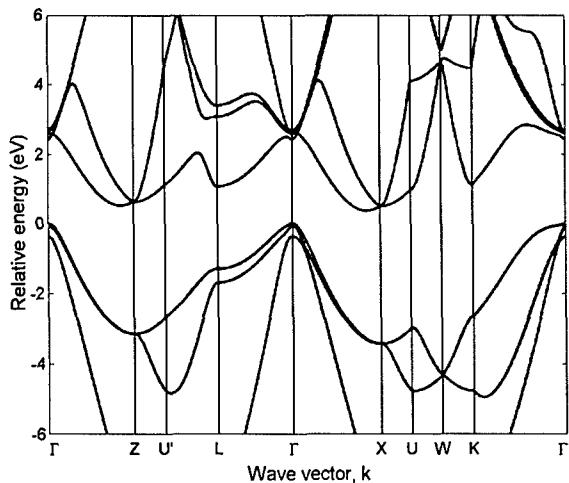


Figure 1. Band structure of strained- $\text{Si}_{1-x}\text{Ge}_x$, $x = 40\%$.

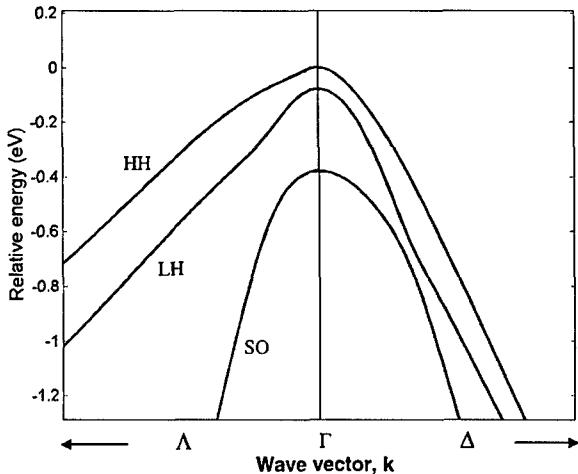


Figure 2. Zoom-in of valence band maximum in Fig. 1.

structure is the splitting of the heavy hole (HH) and light hole (LH) bands at the valence band maximum, which is located at the Γ point (Fig. 1). At $x = 40\%$, the splitting is calculated to be approximately 80 meV. Furthermore, strain also serves to warp the valence bands near the Γ -point. In addition, the spin splitting is enhanced with strain. The value indicated in Fig. 2 is approximately 400 meV, which is larger than that of pure Ge ($\Delta_{SO} = 300$ meV).

7. Conclusion

In summary, the band structure for strained- $\text{Si}_{1-x}\text{Ge}_x$ was calculated using the empirical pseudopotential

method within the virtual crystal approximation. Alloy information is included into the calculation via the Löwdin quasi-degenerate perturbation theory, and strain is included via elastic theory. Strain serves to split the degeneracy of the HH and LH bands at the Γ -point. Furthermore, band warping results from the strain. Finally, applying the Löwdin quasi-degenerate perturbation theory serves to reduce the computation time by a factor of 10.

Acknowledgments

The authors wish to acknowledge the Office of Naval Research and the National Science Foundation for support of this research.

References

- Chelikowsky J.R. and Cohen M.L. 1974. Phys. Rev. B 10: 12.
Cohen M.L. and Bergstresser T.K. 1966. Phys. Rev. 141: 789.
Cressler J.D. 1995. IEEE Spectrum 32: 49.
Harame D.L., Comfort J.H., Cressler J.D., Crabbé E.F., Sun J.Y.-C., Meyerson B.S., and Tice T. 1995a. IEEE Trans. Electron Devices 42: 455.
Harame D.L., Comfort J.H., Cressler J.D., Crabbé E.F., Sun J.Y.-C., Meyerson B.S., and Tice T. 1995b. IEEE Trans. Electron Devices 42: 469.
Iyer S.S., Patton G.L., Stork J.M.C., Meyerson B.S., and Harame D.L. 1989. IEEE Trans. Electron Devices 36: 2043.
Löwdin P. 1951. J. Chem. Phys. 19: 1396.
Phillips J.C. and Kleinman L. 1959. Phys. Rev. 116: 287.
Rieger M.M. and Vogl P. 1993. Phys. Rev. B 48: 14276.
Saravia L.R. and Brust D. 1968. Phys. Rev. 176: 915.
Vegard L. 1921. Z. Phys. 5: 17.



A Computational Exploration of Lateral Channel Engineering to Enhance MOSFET Performance

JING GUO, ZHIBIN REN AND MARK LUNDSTROM

School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

Abstract. Techniques to engineer a MOSFET's channel in the lateral direction have been proposed to enhance the device performance. In this paper, we present a thorough simulation study to evaluate the feasibility of such lateral engineering techniques. Each of three types of transport equations, the ballistic Boltzmann, drift-diffusion and non-equilibrium Green's function with scattering, is solved self-consistently with 2-D Poisson equation to simulate device performance under both the ballistic and dissipative transport conditions. The results indicate that even if highly idealized device structures are assumed, only limited improvements over the conventional MOSFETs can be achieved by the channel engineering techniques. These results don't conflict with reports of large on-current improvements using the lateral channel engineering, because those comparisons with the conventional MOSFETs were done without specifying a common off-current.

Keywords: lateral channel engineering, hetero-material gate MOSFETs, ballistic transport, Green's function

1. Introduction

The success of the microelectronics industry has kept the channel length of MOSFETs scaling down by a factor of 70% about every three years over the past decades. ITRS target for the on-current remains the same while that for the off-current doubles from generation to generation (SIA, 1999), which suggests that as the device scales down, the on-current to off-current ratio, I_{on}/I_{off} , decreases. Since larger I_{on}/I_{off} can provide faster speed and lower leakage, designing MOSFETs with enhanced I_{on}/I_{off} is of wide interest. Techniques to engineer the channel in the lateral direction to improve MOSFET performance have been proposed. The main consideration of such designs is to produce a desired profile along the channel direction as shown in Fig. 1, compared with that of the conventional MOSFETs. The modified profile can generate a larger electric field at the beginning of the channel and results in a larger carrier injection velocity, which is supposed to increase I_{on} . The potential profile can be generated by either using a gate with dual work-functions, which is referred as the hetero-material gate MOSFET (HMGFET) (Long and Chin 1997, Zhou and

Long 1998, Zhou 2000), or doping the source end of the channel more heavily than the rest, which is referred as the asymmetric channel doping MOSFETs (ACDFETs) (Odanaka and Hiroki 1997, Shin and Lee 1999). Another lateral channel engineered device, the straddle gate MOSFET (Tiwari, Welser and Solomon 1998), which uses two side gate beside the inner gate with a different work function, is based on the consideration to electrically reduce the effective channel length from the off-state to the on-state. In this paper, we compare lateral-channel engineered MOSFETs and conventional MOSFETs with the same off-current and geometric specifications under both the ballistic and dissipative transport conditions. The results indicate that even if highly idealized device parameters are assumed, only limited improvements can be achieved.

2. Approach

The 2-D transport equation in the MOSFET channel region is solved by splitting it into two 1-D problems. In the direction normal to the channel, the Schrodinger equation is solved to yield subband profile and vertical

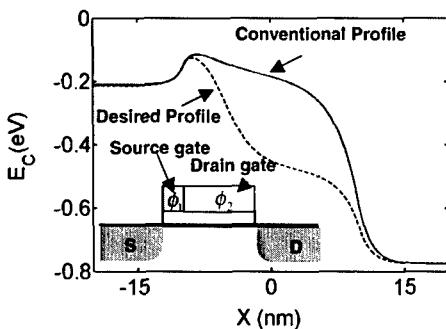


Figure 1. The conventional (solid line) and desired (dashed line) potential profile along the lateral direction of the channel. Also shown is the schematic structure of HMGFETs which can generate the desired potential profile if the source gate workfunction ϕ_1 is larger than the drain gate workfunction ϕ_2 .

electron concentration. In the lateral direction, three types of transport equations, the Boltzmann equation in the ballistic limit, drift-diffusion equation and Non-Equilibrium Green's Function (NEGF) with scattering are solved to yield electron density in the lateral direction and the source-drain current on the basis of the subband profiles. A 2D Poisson equation is solved self-consistently with each of the transport equations. Details of calculation scheme can be found in Taur and Ning (1998).

The ballistic limit is calculated semiclassically by solving Boltzmann equation. For each spatial point, the occupation of a state in k -space is determined by the Fermi-Dirac function with the source or drain Fermi level, depending on which contact the electrons fill such state come from. For example, at the barrier top, the positive half of k -space is filled by electrons from the source while the negative half by electrons from the drain.

In our Green's function method, we treat scattering using a simple Buttiker-probe model. Scattering centers are viewed as reservoirs similar to the source and drain. However, they differ from the source and drain reservoirs as they can only change the energy of carriers and not the total number of carriers in the system. This model has been demonstrated to capture the essential physics of scattering (Taur and Ning 1998).

3. Results and Discussions

The characteristics of a conventional double-gate MOSFET at 30-nm channel length, as shown in Fig. 2(a), are calculated as the comparison baseline to

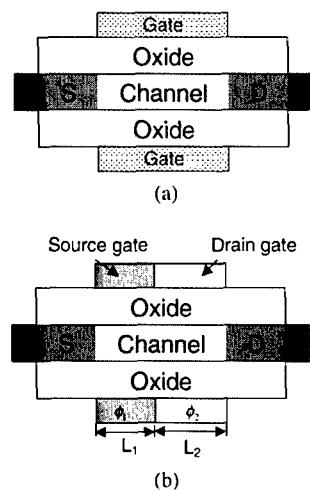


Figure 2. (a) A conventional symmetric double-gate MOSFET with 30-nm intrinsic channel. An ultrathin body and oxide $t_{ox} = 2$ nm, $t_{ni} = 1$ nm and a middle gap gate workfunction $\phi_G = 4.3$ V are assumed. The top and bottom gates have equal lengths with the channel. (b) HMGFET with the same geometric dimensions. The total gate length is the sum of the source gate length L_1 and the drain gate length L_2 .

evaluate the performance of HMGFETs. An ultrathin silicon body and gate oxide are assumed in order to suppress 2-D short channel effects. While 2 nm Si body thickness and 1 nm gate oxide thickness can hardly be achieved with current fabrication technologies and may cause problems such as the gate leakage, the purpose of exploiting such parameters is to evaluate the maximum achievable improvement of HMGFETs under highly idealized conditions. Our comparisons between HMGFETs and the conventional MOSFET are done by specifying the common geometric parameters and off-current. For HMGFETs as shown in Fig. 2(b), the off-current is kept the same with that of the conventional MOSFET by choosing an appropriate source gate workfunction, ϕ_1 , and assuming the drain gate workfunction, $\phi_2 = 4.05$ V. The characteristics are then simulated and compared with the corresponding ones of the conventional MOSFET.

To determine the source gate length L_1 , the first subband profiles of HMGFETs with different L_1 values at on-state are simulated as shown in Fig. 3. HMGFETs with short source gate length generate preferable potential profile in the consideration of maximizing the electric field near the subband barrier top. Increasing the source gate length results in the decrease of the electric field near the source, which makes the subband profile approach that of the conventional MOSFET. When the source gate length is longer than one half of

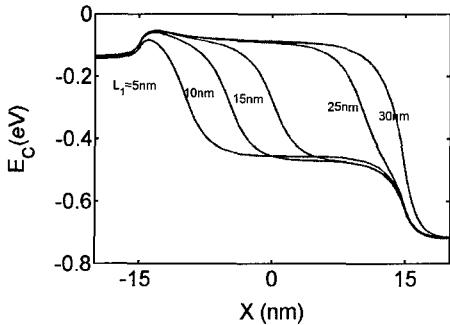


Figure 3. The potential profile along channel at the on-state for HMGFETs with different L_1 . $\phi_1 = 4.4$ V and $\phi_2 = 4.05$ V are kept constant.

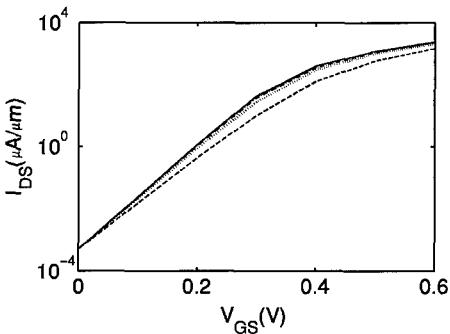


Figure 4. The I_{DS} vs. V_{GS} characteristics at $V_D = 0.6$ V calculated by the ballistic transport model. Solid line: the conventional MOSFET shown in Fig. 2. Dash line: $L_1 = 5$ nm HMGFET with $\phi_1 = 4.46$ V. Dot line: $L_1 = 10$ nm HMGFET with $\phi_1 = 4.33$ V. Dash-dot line: $L_1 = 15$ nm HMGFET with $\phi_1 = 4.31$ V.

the total gate length, the subband profile near the barrier top is almost the same as that of the conventional MOSFET. HMGFETs with three different source gate lengths $L_1 = 5$ nm, 10 nm and 15 nm are studied in the subsequent ballistic and dissipative transport calculations.

Figure 4(a) shows I_{DS} - V_{GS} characteristics of HMGFETs, compared with that of the conventional MOSFET. The $L_1 = 5$ nm HMGFET has a larger subthreshold swing and worse short-channel immunity. Increasing the source gate length to 10 and 15 nm improves the subthreshold characteristics, which are mainly dominated by the device electrostatics. The degraded electrostatic properties associated with HMGFETs can be understood by qualitatively analyzing 2-D Poisson equation in the channel region. The slope of subband profile of HMGFETs increases from zero at the barrier top to a large value in order to increase carrier injection velocity, indicating a rapidly spatial change of electric field at the position. The

Table 1. On-current evaluation of three HMGFETs using different transport models, where the units of I_{on} are $\mu\text{A}/\mu\text{m}$ and Increase indicates the percentage increase of I_{on} over the conventional MOSFET with the same transport model.

	Ballistic		DD		NEGF	
	I_{on}	Increase (%)	I_{on}	Increase (%)	I_{on}	Increase (%)
Conventional	2340	—	810	—	1378	—
$L_1 = 5$ nm HMG	1439	-38	605	-25	1248	-10
$L_1 = 10$ nm	2060	-12	829	+2	1615	+17
$L_1 = 15$ nm	2262	-3	882	+10	1643	+20

rapidly varying field results in the large absolute value of the second derivative of the potential along the channel, which may invalidate the gradual channel approximation and lead to more severe 2-D short channel effects (Ren 2001). Such short channel effects cause less effective gate modulation on the barrier top of HMGFETs, especially for those HMGFETs with short source gate length. Thus the barrier top of HMGFETs cannot be pushed down so much as that of the conventional MOSFET from the off-state to the on-state. The higher subband barrier tops result in the reduction of the ballistic on-current as indicated in Table 1.

In the presence of scattering, the characteristics of HMGFETs and the conventional MOSFET are first calculated using the drift-diffusion model. From the design consideration of the lateral channel engineering, it might be expected that the largest I_{on} improvement can be achieved by $L_1 = 5$ nm HMGFET because it maximizes the electric field near the barrier top. However, the results indicate an opposite situation as shown in Table 1. Although the $L_1 = 5$ nm HMGFET does achieve the largest carrier injection velocity V_{inj} as shown in Fig. 5(a), the injection carrier density reduction, which is shown in Fig. 5(b) is more dominant and causes the overall decrease of the on-current. Such reduction can be explained on the basis of simple gate control electrostatics, which express the injection carrier density as $Q_{inj} = C_{eff}(V_G - V_T)$, where C_{eff} is the effective gate capacitance, V_G is the gate voltage and V_T is the threshold voltage. The worse subthreshold characteristics of $L_1 = 5$ nm HMGFET requires a larger V_T to yield the specified off-current, thus causing the decrease of Q_{inj} at on-state when the same C_{eff} is assumed. Increasing the source gate length can lead to larger Q_{inj} , however, at the same time, it decreases V_{inj} as shown in Fig. 5. This trade-off relation between Q_{inj} and V_{inj} makes it hard to achieve large on-current improvement. In the

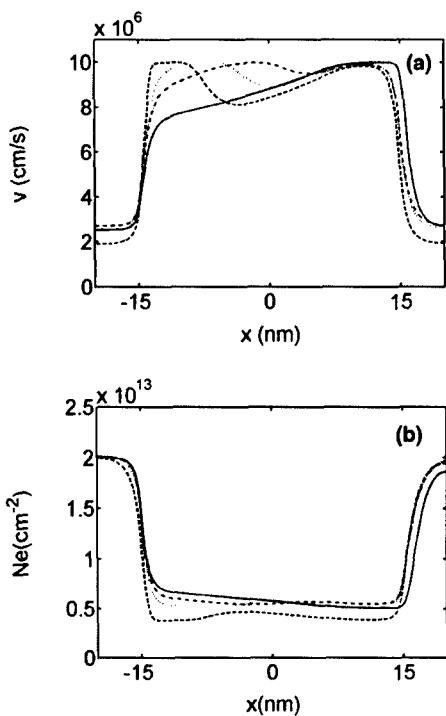


Figure 5. (a) The velocity distribution and (b) the electron density along the lateral direction at on-state calculated using drift-diffusion model for MOSFETs with the same symbols as Fig. 4.

best case when $L_1 = 15$ nm, a maximum improvement of about 10% is obtained, which is shown in Table 1.

Drift-diffusion treatment misses transport mechanisms such as quantum tunneling and velocity overshoot, which can be important for small dimensions. To include physics beyond DD model, the NEGF approach with scattering was employed to recalculate the device characteristics at the on-state. A typical velocity distribution curve of HMGFET as shown in Fig. 6

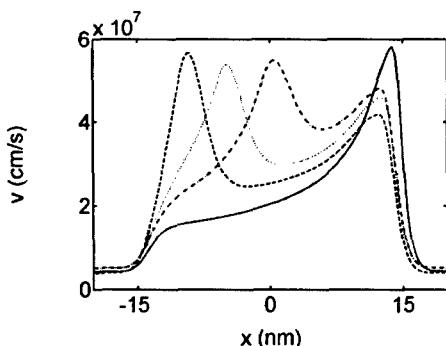


Figure 6. The velocity distribution at on-state calculated using NEGF. For MOSFETs with the same symbols as Fig. 4.

displays two overshoot peaks, one at the boundary between the source and drain gate, the other near the drain end. These two peaks are related to the rapidly spatially increase of electric field at these two regions. The left velocity overshoot peak can yield larger improvement of carrier injection velocity than the drift-diffusion model results, thus corresponding to larger on-current improvement as shown in Table 1. About 20% improvement was attained in the best case.

Since asymmetric channel doping is essentially based on the same design consideration as HMGFETs, similar observations apply to such device. The present study uses double-gate structure with extremely thin gate oxide and Si body thickness to suppress the short channel effects. If more realistic parameters are used and non-ideal conditions, such as parasitic resistance, included, the improvement achievable by using lateral channel engineering would become even smaller.

It is also worth pointing out that our results don't contradict most of the reported large improvements of on-current achieved by lateral channel engineering because these comparisons were done without specifying a common off-current. Comparing on-currents without considering the off-state or by specifying a common threshold voltage can leave the worse subthreshold performance of the lateral channel engineered MOSFETs out and lead to larger improvements. One exception reporting better subthreshold performance by exploiting hetero-material gate structure needs further study (Long and Chin 1997).

4. Conclusions

Each of three types of transport equations is solved self-consistently with 2-D Poisson equation to compare the performance of conventional MOSFETs and HMGFETs under both the ballistic and dissipative transport conditions. The ballistic results indicate that HMGFETs have larger subthreshold swing and threshold voltage than the conventional MOSFETs due to the short channel effects, leading to smaller on-current. After including scattering, we showed that the higher carrier injection velocity of HMGFETs due to larger electric field near the barrier top doesn't necessarily lead to larger on-current. For many cases, the lateral field gradient degrades shot channel performance, so for a specified off-current, the threshold voltage is higher, which makes the on-current smaller. When highly idealized device parameters are used, a maximum improvement of 10–20% can be achieved. Such observations can also

be extended to ACDFETs, which is essentially based on the similar design consideration as HMGFETs.

Acknowledgment

This work was supported by Semiconductor Research Corporation under the contract number 99-NJ-724.

References

- Long W. and Chin K.K. 1997. IEDM Tech Digest. IEEE, Washington DC, p. 549.
- Odanaka S. and Hiroki A. 1997. IEEE Trans Electron Devices 44: 595.
- Ren, Z. 2001. Ph.D. Thesis. Purdue University, West Lafayette, IN, USA. Unpublished.
- Semiconductor Industry Association (SIA). 1999. International Technology Roadmap for Semiconductors.
- Shin H. and Lee S. 1999. IEEE Trans. Electron Devices 46: 820.
- Taur Y. and Ning T.H. 1998. Fundamentals of Modern VLSI Devices. Cambridge University Press, Cambridge, UK, p. 144.
- Tiwari S. Welser J.J. and Solomon P.M. 1998. IEDM Tech Digest. IEEE, San Francisco, p. 737.
- Zhou X. 2000. IEEE Trans. Electron Devices 47: 113.
- Zhou X. and Long W. 1998. IEEE Trans. Electron Devices 45: 2546.



Monte Carlo Simulations of Hole Dynamics in Si/SiGe Quantum Cascade Structures

Z. IKONIĆ, P. HARRISON AND R.W. KELSALL

*Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering,
University of Leeds, Leeds LS2 9JT, UK*

r.w.kelsall@leeds.ac.uk

Abstract. We report the first detailed ensemble Monte Carlo simulation of hole dynamics in cascaded *p*-Si/SiGe quantum wells. The hole subband structure is calculated using the $6 \times 6 \mathbf{k} \cdot \mathbf{p}$ model. The simulation accounts for the in-plane \mathbf{k} -space anisotropy of both the hole subband structure and the scattering rates. The scattering mechanisms included are the alloy disorder, acoustic and optical phonon scattering. Results are presented for prototype Si/SiGe cascade structures.

1. Introduction

There has recently been an increased interest in intersubband transitions in *p*-type strained-layer SiGe based quantum wells, due to their possible use in intersubband quantum cascade lasers operating in the mid-to far-infrared wavelength range (Soref, Friedman and Sun 1998, Friedman *et al.* 1998). This is largely related to the fact that hole intersubband transitions are optically active for both the perpendicular and the in-plane polarization of light, hence enabling the realization of surface emitting intersubband lasers. These points, together with the comparatively small cost of SiGe, as compared to III-V based structures, and the possibility of monolithic integration of electronic and optoelectronic components based on this system are strong incentives for the development of a SiGe cascade laser.

Understanding the carrier dynamics in cascade lasers is an important issue for the design of these structures. The gain depends sensitively on the scattering rates between different subbands and also between different in-plane momentum states within a subband (carrier heating/cooling effects). In lasers based on conduction band intersubband transitions the gain may be reasonably accurately estimated within the self-consistent rate equation model (Donovan, Harrison and Kelsall 2001).

In lasers based upon transitions among the valence band states, however, there is an added complexity in that the hole scattering rates and the optical transition matrix elements are anisotropic and strongly dependent on the in-plane momentum of hole states. This situation requires a more detailed approach, such as that provided by the Monte Carlo (MC) method. Here we describe the implementation of the MC method for calculating holes dynamics in cascaded SiGe structures.

2. Calculation Details

The MC method has a long history of successful applications in modelling carrier dynamics in semiconductors (Jacoboni and Reggiani 1983), and has been used in calculations of properties such as electron and hole mobility at high and low fields and impact ionization in both bulk semiconductors and in 2- and 3-terminal devices. The method has also been used to study carrier relaxation processes in low-dimensional structures (quantum wells) (Dür, Goodnick and Lugli 1996). Quite recently, MC simulation of lasers based on electronic (conduction band) intersubband transitions has been performed, and a great deal of insight in the electron dynamics in both optically pumped (Kelsall, Kinsler and Harrison 2000) and electrically pumped

(quantum cascade) (Iotti and Rossi 2001a, b) devices was gained. MC studies of hole dynamics in quantum wells (in-plane mobility) have also been reported (Kelsall *et al.* 1992). This problem is generally similar, though somewhat more complex, than the electron case. This is because of the presence of different types of holes (heavy, light, and split-off), which gives rise to mixing of these bulk states in the quantized states of the system, and in turn results in prominent in-plane anisotropy and nonparabolicity of hole subbands. Furthermore, all the scattering processes, and the optical transitions between the quantized states, exhibit both anisotropy and in-plane momentum dependence.

The MC calculation developed in this work uses hole bandstructure data precalculated using the $6 \times 6 \mathbf{k} \cdot \mathbf{p}$ scheme (Foreman 1993, Ikonić, Harrison and Kelsall in press). The energies and wavefunctions of the subbands of interest are tabulated at a number of in-plane \mathbf{k} states in the irreducible wedge of the 2D Brillouin zone (for the usual, [001] grown structures this is 1/8 of the full 2D Brillouin zone), and, due to the symmetry this is sufficient to account for the full anisotropy of the band structure, which has been found previously (Ikonić, Harrison and Kelsall in press) to be important in the scattering rate calculation. This data is then used by the MC code to find the microscopic (differential) scattering rates between all pairs of states, including both intrasubband and intersubband transitions. This is accomplished using a cellular scheme, in which the 2D Brillouin zone is subdivided into a grid of phase space cells, and the scattering rate from each cell into any other cell is calculated. The results are stored in a look-up table, to be used in the main part of the MC code. The scattering processes currently included are alloy disorder, acoustic phonon, and optical phonon scattering (the latter includes Ge-Ge, Ge-Si, and Si-Si modes). In the case of acoustic phonon scattering, the linear dispersion of phonons is included, because it was found to have a significant effect on the scattering lifetimes (Ikonić, Harrison and Kelsall in press). Each type of scattering process has a separate entry in the look-up table.

The MC code works with a constant timestep (Goodnick and Lugli 1988, Fischetti and Laux 1988), which is determined initially by inspection of the look-up table. This is less common than the standard approach, where the time elapsed between scatterings is generated randomly. The main advantage of this approach is that the hole ensemble always stays synchronised in time, which simplifies the simulation

of the Pauli exclusion effect, and also of carrier-carrier scattering, although this is achieved at the cost of an increased number of self-scatterings. When tracking the hole dynamics, interpolation is used to construct a sub-table of scattering rates from the particular hole state (with the actual value of \mathbf{k}) into other states (cells). After assembling its entries, multiplied by the timestep, into a table of accumulated scattering probabilities, a random number is generated and ranked in this table, wherefrom it is decided whether the particular event is a real scattering or a self-scattering. If it is a real scattering, the ranking simultaneously decides not only the cell that the final state belongs to, but also the type of scattering that occurred. Furthermore, the angular dependence of the scattering probability, which is a separate phase in the conventional approach, is implicitly contained in the look-up table. At that stage the Pauli exclusion based acceptance or rejection of this event is applied, in the manner described in Lugli and Ferry (1985). If accepted, the precise \mathbf{k} of the final state is found by generating \mathbf{k} values at random within the final cell, and testing for energy conservation until this is satisfied and a state is finally accepted.

Additional considerations are necessary when MC simulation of a quantum cascade structure is required. Our approach involves applying periodic boundary conditions, which impose the condition that the particle distribution in each period of the structure is identical in the steady state. In the prototype *p*-SiGe cascades we have considered, each period comprises a single SiGe quantum well. In MC simulations of such structures we assume that only transitions between neighbouring wells are important (“nearest neighbour interaction”) (Iotti and Rossi 2001b). This is justifiable in *p*-SiGe cascades because the hole wavefunctions are each strongly localized within a single well. To be able to track the hole dynamics in a cascade structure it is sufficient to calculate a table of microscopic scattering rates for a structure comprising just two coupled wells in the presence of a uniform electric field, since this table will contain the rates for all interwell (both “upstream” and “downstream” nearest neighbour) transitions in a periodic cascade structure, as well as all intrawell transitions. In the actual MC simulation a section of the cascade with three wells is considered, in which the initial hole state can only be one of the central well states, while the final state of a scattering process can be in any of the three wells. The transition probabilities are read from the two-well

scattering rate table, and when the final state in the three-well system is chosen it is mapped back into the central well (this is the implementation of the periodic boundary condition). Interwell scatterings are current-carrying, in either the upstream or downstream directions.

3. Numerical Results and Discussion

A set of MC simulations has been performed for several *p*-SiGe cascade structures. For the band structure calculation, the material parameters for Si and Ge were taken from Kahan, Chi and Friedman (1994), and the valence band edge discontinuity from Van de Walle and Martin (1986). The phonon and alloy scattering parameters have been taken from Crow and Abram (2000) and Kearney and Horrell (1998).

The first structure considered was a coupled well system with 20 monolayer (55 Å) $\text{Ge}_{0.4}\text{Si}_{0.6}$ wells and 20 monolayer Si barriers, grown on a $\text{Ge}_{0.2}\text{Si}_{0.8}$ virtual substrate, biased at 85 kV/cm, as shown in Fig. 1. The HH2-LH1 and LH1-HH1 spacings are 34 meV and 52 meV respectively. Figure 2 shows the result of a transient MC simulation for this structure, following the population of states after pulsed injection into the HH2 subband of the left well. For the equilibrium initial hole population (within the left HH2 subband) we find a generally good agreement between the relaxation times extracted from MC simulation and those evaluated in the conventional manner. A periodic MC simulation of the cascade structure was also performed, and indicated that this structure, at a lattice temperature of $T = 20$ K, is on the verge of achieving population inversion between the HH2 and LH1 states, but has a drawback in that the majority

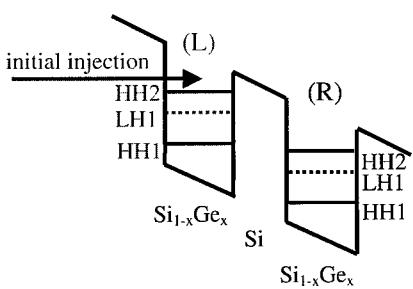


Figure 1. Schematic diagram of a coupled quantum well structure used in MC simulations.

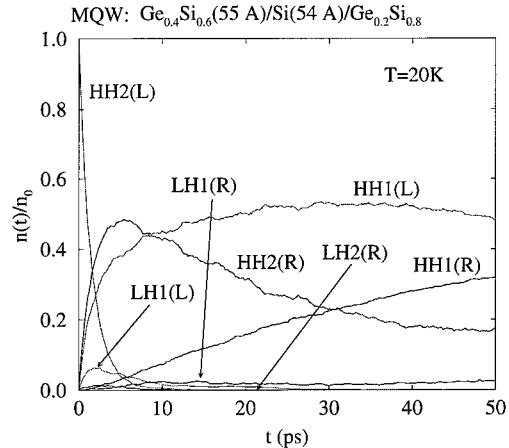


Figure 2. Time dependent populations in the structure from Fig. 1, following pulse injection into the HH2(L) subband.

of holes in the stationary state reside idle in the HH1 state.

The next set of cascade structures considered has 16 monolayer $\text{Ge}_{0.3}\text{Si}_{0.7}$ wells and thin Si barriers (4, 6, or 8 monolayers), grown on a $\text{Ge}_{0.2}\text{Si}_{0.8}$ virtual substrate. The wells have just two subbands, the ground HH1 and the excited LH1, which are reasonably low in energy (accessible to holes under <100 kV/cm bias). The bias field is large enough that the HH1 state in the preceding well is above the LH1 state in the next well. The spacing between these two states depends on the bias, while the intra-well LH1-HH1 spacing is essentially fixed by the structural parameters to ≈ 30 meV, and is almost independent of bias. It may be interesting as a possible laser structure where the lasing transition would be the inter-well HH1(L) \rightarrow LH1(R), while the intra-well LH1(R) \rightarrow HH1(R) is the relaxation transition, emptying the lower laser state (cf. Fig. 1 for notation). This is because the optical matrix element at the zone center, for the in-plane polarized light, is smaller for the intra- than for the inter-well transition. Results of the periodic MC simulation are shown in Tables 1 and 2. Concerning the total occupancy of subbands, the agreement between the MC and rate equations results is reasonable for wider barriers, but significant discrepancies appear for thin barriers/high field structures. These occur because of carrier heating to well above the lattice temperature, which translates into changes of carrier dynamics. On the other hand, the current calculated from the rate equations model is always significantly lower than the MC result, because heated carriers carry disproportionately large fraction of the total current.

Table 1. Steady state periodic Monte Carlo (superscript 'MC') and rate equation (superscript 'r') results for subband populations in a SiGe cascade structure: 16 ml Ge_{0.3}Si_{0.7} wells/8 ml Si barriers, at $T = 77$ K, under different bias fields E . Carrier density was 10^{11} cm⁻² per well. Also given are the transition energy E_{tr} and the optical transition matrix element $p_{opt}^2/2m_0$ for the in-plane polarization.

E (kV/cm)	n_{HH1}^{MC}	n_{LH1}^{MC}	I^{MC} (A/cm ²)	n_{HH1}^r	n_{LH1}^r	I^r (A/cm ²)	E_{tr} (meV)	$p_{opt}^2/2m_0$ (meV)
60	71%	29%	245	71%	29%	133	13	1.03
80	89%	11%	257	92%	8%	68	25	1.05
100	90%	10%	406	94%	6%	105	39	1.20

Table 2. Same as Table 1, but for 16 ml Ge_{0.3}Si_{0.7} wells/(4–8) ml Si barriers, under bias fields adjusted to maintain $E_{tr} = 25$ meV transition energy.

w_{Si} (ml)	E (kV/cm)	n_{HH1}^{MC}	n_{LH1}^{MC}	I^{MC} (A/cm ²)	n_{HH1}^r	n_{LH1}^r	I^r (A/cm ²)	$p_{opt}^2/2m_0$ (meV)
4	94	76%	24%	880	85%	15%	208	3.80
6	84	83%	17%	460	88%	12%	114	2.10
8	80	89%	11%	257	92%	8%	68	1.05

4. Conclusion

Detailed ensemble MC simulations of hole dynamics in cascaded Si/SiGe quantum wells were performed, with the in-plane \mathbf{k} -space anisotropy of both the hole subband structure and the scattering rates taken into account. The results indicate the value of the MC simulation method for reliable calculations of carrier dynamics in quantum cascade structures.

Acknowledgments

This work is supported by DARPA/USAF contract No. F19628-99-C-0074. The authors thank R.A. Soref (Hanscom AFB) for useful discussions.

References

- Crow G.C. and Abram R.A. 2000. Semicond. Sci. Technol. 15: 7.
- Donovan K., Harrison P., and Kelsall R.W. 2001. J. Appl. Phys. 89: 3084.
- Dür M., Goodnick S.M., and Lugli P. 1996. Phys. Rev. B 54: 17794.
- Fischetti M.V. and Laux S.E. 1988. Phys. Rev. B 38: 9721.
- Foreman B.A. 1993. Phys. Rev. B 48: 4964.
- Friedman L., Soref R.A., Sun G., and Lu Y. 1998. IEEE J. Select. Topics Quantum Electron. 4: 1029.
- Goodnick S.M. and Lugli P. 1988. Phys. Rev. B 37: 2578.
- Ikonić Z., Harrison P., and Kelsall R.W. Phys. Rev. B, in press.
- Iotti R.C. and Rossi F. 2001a. Appl. Phys. Lett. 78: 2902.
- Iotti R.C. and Rossi F. 2001b. Phys. Rev. Lett. 87, Art No. 146603.
- Jacoboni C. and Reggiani L. 1983. Rev. Modern Phys. 55: 645.
- Kahan A., Chi M., and Friedman L. 1994. J. Appl. Phys. 75: 8012.
- Kearney M.J. and Horrell A.I. 1998. Semicond. Sci. Technol. 13: 174.
- Kelsall R.W., Abram R.A., Batty W., and O'Reilly E.P. 1992. Semicond. Sci. Technol. 7: 86.
- Kelsall R.W., Kinsler P., and Harrison P. 2000. Physica E 7: 48–51.
- Lugli P. and Ferry D.K. 1985. IEEE Trans. Electron Devices ED-32: 2431.
- Soref R.A., Friedman L., and Sun G. 1998. Superlatt. Microstr. 23: 427.
- Van de Walle C.G. and Martin R.M. 1986. Phys. Rev. B 34: 5621.



Calculation of Direct Tunneling Current through Ultra-Thin Gate Oxides Using Complex Band Models For SiO₂

ATSUSHI SAKAI, AKIHIRO ISHIDA, SHIGEYASU UNO AND YOSHINARI KAMAKURA

Department of Electronics and Information Systems, Osaka University, Osaka 565-0871, Japan

MASATO MORIFUJI

Department of Electronic Engineering, Osaka University, 2-1 Yamada-oka, Suita, Osaka 565-0871, Japan

KENJI TANIGUCHI

Department of Electronics and Information Systems, Osaka University, Osaka 565-0871, Japan

Abstract. We report the calculation of gate leakage currents through the ultra-thin gate oxides (2.6–3.4 nm) in MOSFETs. We simulate *J-V* characteristics for the direct tunneling of valence electrons and inversion layer holes, which are measured using a charge separation technique. A two-band model is employed to express the complex band structure of the gate oxide, and its validity is discussed by calculating the complex band structure of β -cristobalite based on the second nearest neighbor sp^3s^* tight-binding scheme.

Keywords: gate leakage current, direct tunneling, MOSFET, two-band model, complex band structure, tight-binding scheme

1. Introduction

Understanding the mechanism of gate leakage current due to direct tunneling (DT) is essential for modeling the operation of Si-MOSFETs with gate oxide thicknesses below 3 nm. It has been reported that DT current through the ultra-thin oxide layer consists of various components: DT of conduction band electrons in the electrode, DT of valence electrons, DT of holes in the inversion layer, etc. Many parameter sets (the tunneling effective mass and the barrier height energy) have been determined to model each DT component (Lee and Hu 2001).

The goal of this study is the unified understanding of various DT currents. We report in the present paper the calculation of *J-V* characteristics for DT currents carried by valence electrons and inversion layer holes, which is measured by a charge separation technique. We use a two-band model (Kane 1966) to describe the complex band structure in the whole range of SiO₂ band gap. In addition, we calculate the

complex band structure of SiO₂ based on the tight-binding model (Schulman and Chang 1983, Ting, Yu and McGill 1992, Boykin 1996, Städele, Tuttle and Hess 2001) and discuss the accuracy of the empirical DT current calculations.

2. Physical Models and Numerical Techniques

2.1. Tunneling Current Model

We considered one-dimensional p^+ -Si(100)/SiO₂/n-Si(100) sandwiches. The self-consistent calculation of Poisson and Schrödinger equations was performed to evaluate the band bending in both substrate and gate depletion layers. The tunneling current was obtained by applying the Tsu-Esaki's formula (Tsu and Esaki 1974) to the cases analyzed in this study. The tunneling from the bound states in the inversion layer was calculated with the method reported by Rana, Tiwari and Buchanan (1996) and Ghetti *et al.* (1999).

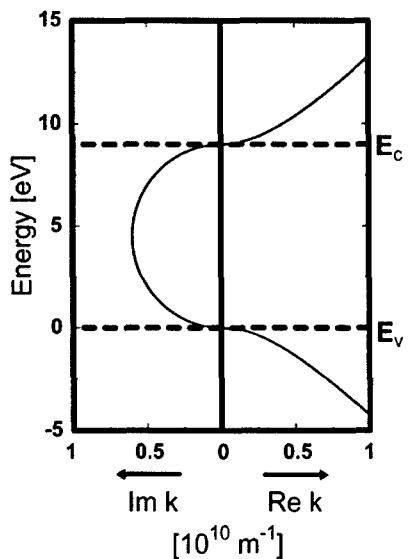


Figure 1. Complex band structure for the gate oxide based on a two-band model ($m_{ox}^* = 0.6m_0$ and $E_g = 9$ eV).

For conduction electrons in silicon a many valley ellipsoidal parabolic band ($m_{lc}^* = 0.916m_0$, $m_{le}^* = 0.19m_0$) was assumed. The silicon valence band was expressed by introducing the effective mass perpendicular to Si/SiO₂ interface, m_{\perp}^* , and the density-of-state mass, m_d^* , for the parallel direction to the interface (Takagi, Takayanagi and Toriumi 1999). The numerical values used were $m_{\perp,hh}^* = 0.29m_0$, $m_{d,hh}^* = 0.433m_0$, $m_{\perp,lh}^* = 0.20m_0$, $m_{d,lh}^* = 0.169m_0$. In order to calculate the transmission probability the information about E - k relationship in SiO₂ band gap, i.e. the complex band structure, is necessary. In this study, we used a two-band model (Kane 1966):

$$k = \sqrt{\frac{2m_{ox}^*}{\hbar^2 E_g}}(E - E_C)(E - E_V), \quad (1)$$

where m_{ox}^* is the effective mass, E_C and E_V are the energies at the bottom of the conduction band and the top of the valence band, respectively, and E_g is the band gap energy. As shown in Fig. 1, the wave number k corresponding to the energy E in the gap region is an imaginary number, which describes the decay of wave function in the barrier.

2.2. Complex Band Calculation Using Tight-Binding Scheme

In order to verify the validity of the two-band model for the complex band structure of SiO₂, we carried out

Table 1. Empirical tight-binding parameters for the sp^3s^* model in eV. The notation is that of LaFemina and Duke (1991) and Vögl, Hjalmarson and Dow (1983).

$E_s[Si]$	14.260
$E_p[Si]$	18.360
$E_{s^*}[Si]$	39.455
$E_s[O]$	-16.360
$E_p[O]$	-1.770
$E_{s^*}[O]$	20.270
$V_{ss\sigma}[Si - O]$	-1.500
$V_{ss\sigma}[O - O]$	0.250
$V_{sp\sigma}[Si - O]$	3.046
$V_{sp\sigma}[O - Si]$	-3.760
$V_{sp\sigma}[O - O]$	-0.075
$V_{pp\sigma}[Si - O]$	5.710
$V_{pp\sigma}[O - O]$	1.290
$V_{s^*p\sigma}[Si - O]$	1.016
$V_{s^*p\sigma}[O - Si]$	-1.560
$V_{s^*p\sigma}[O - O]$	0.080
$V_{s^*s^*\sigma}[Si - O]$	-6.700
$V_{s^*s^*\sigma}[O - O]$	0.000

the tight-binding calculation. Although the gate oxide in MOSFETs is an amorphous SiO₂, we assumed a β -cristobalite structure (Gnani *et al.* 2000) for simplicity of computation. We modified the second nearest neighbor sp^3 model for the bulk β -cristobalite reported in LaFemina and Duke (1991); we add an excited s state to the basis (Vögl, Hjalmarson and Dow 1983) and readjusted the parameters to reproduce effective mass of $0.5m_0$ in [100] direction at the bottom of the conduction band (Gnani *et al.* 2000). The resulting effective mass at the top of the valence band is $2.1m_0$. The tight-binding parameters are given in Table 1. With this parameter set, we calculate the complex band structure of SiO₂ by using the technique reported in Schulman and Chang (1983), Ting, Yu and McGill (1992) and Boykin (1996).

3. Experimental

The samples used were p^+ -polysilicon-gate MOSFETs fabricated on (100) oriented n -Si substrates. The oxide thicknesses were 2.6, 3.0, and 3.4 nm, which were measured by ellipsometric technique. The doping concentrations in the gate and the substrate were 6.5×10^{19} cm⁻³ and 4.5×10^{17} cm⁻³, respectively. Figure 2

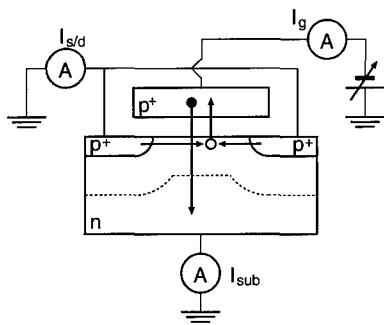


Figure 2. Experimental arrangement for charge separation measurement.

shows the experimental setup for the charge separation technique. Two gate current components can be separately measured; the hole tunneling current from the inversion layer is detected at the source/drain electrodes, and the valence electrons injected from the gate into the substrate are collected at the substrate electrode.

4. Results and Discussions

Figure 3 shows the results of measured and calculated DT currents as a function of the gate voltage. Here the oxide thickness was used as a fitting parameter because there exists uncertainty in the data of ellipsometric measurement (Ghetti *et al.* 1999). The oxide

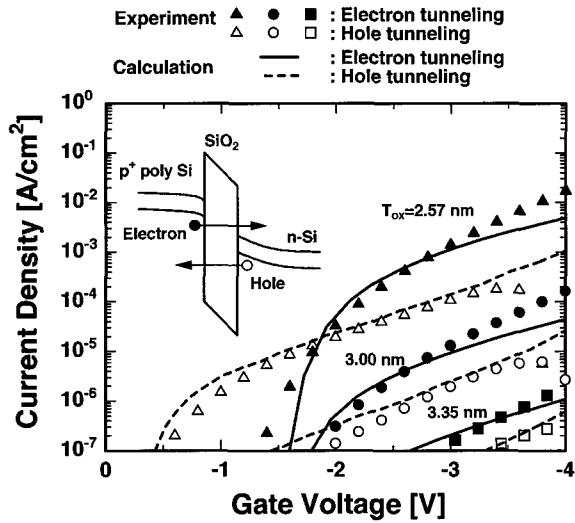


Figure 3. Measured (symbols) and calculated (lines) tunneling currents through the gate oxide layer in p^+ poly pMOSFETs as a function of the gate voltage. The oxide thicknesses (in parenthesis the corresponding value from ellipsometric measurements) are 2.57 (2.6), 3.00 (3.0), and 3.35 (3.4) nm.

thicknesses needed to fit experiments are within 5% of the ellipsometric measurements. For both two DT components due to the valence electrons and the inversion layer holes, we obtained fairly good agreements with the experimental data by using the two-band model with $m_{\text{ox}}^* = 0.6m_0$.

Figures 4–6 show the complex band structures obtained by the tight-binding scheme for the

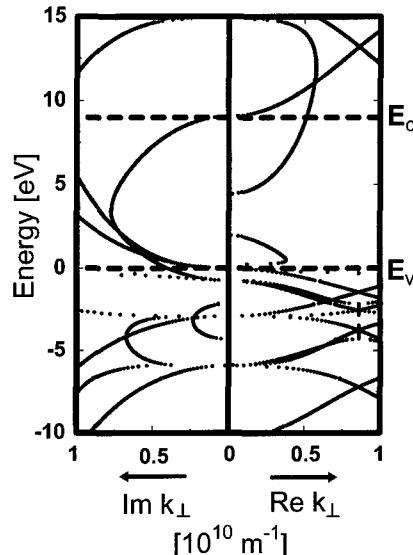


Figure 4. Complex band structure of SiO_2 along [100] direction obtained by the tight-binding scheme.

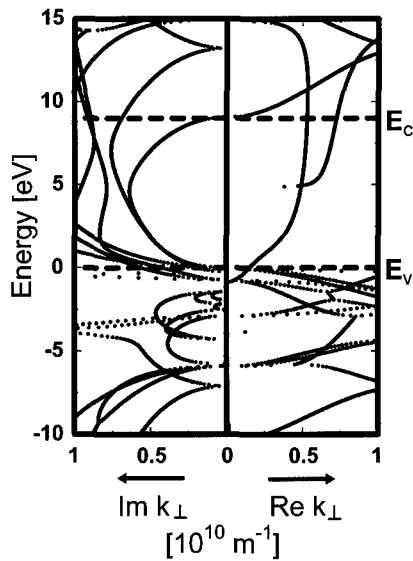


Figure 5. Complex band structure of SiO_2 along [110] direction obtained by the tight-binding scheme.

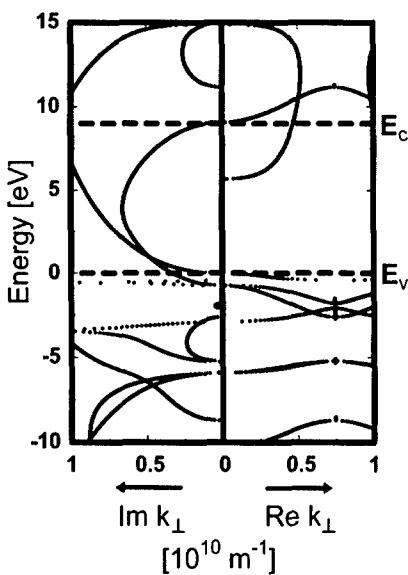


Figure 6. Complex band structure of SiO_2 along [111] direction obtained by the tight-binding scheme.

β -cristobalite along [100], [110], and [111] directions, respectively. Because we have insufficient knowledge about the atomic structure of the gate oxide, we compared three cases to investigate how the atomic configuration affects the results qualitatively. The similar shapes are observed regardless of the wave propagation directions; the semicircular loop connects the conduction and valence bands in the gap region. This property is also found in Städele's tight-binding calculation (Städele, Tuttle and Hess 2001), in which SiO_2 models based on tridymite and β -quartz were used. As shown in Fig. 1, the two-band model expresses this semicircular curve. On the other hand, in the tight-binding results, many other bands with larger $\text{Im } k$ are observed. The complex bands in the gap region with $\text{Re } k \neq 0$ also have large imaginary parts: $\text{Im } k > 1 \times 10^{10} \text{ m}^{-1}$. Their contributions to the tunneling currents are considered to be small, because the larger imaginary k corresponds to the faster decaying wave in the barrier. Moreover, several complex bands are found in the lower half region of the band gap. Probably, their effects are averaged and effectively expressed by a single curve of the two-band model after the fitting procedure of m_{ox} . The wave vector in Fig. 1 shows the maximum toward the midgap and is symmetrical with respect to the midgap. For the analysis of the tunneling in metal-narrow-gap-semiconductor-metal tunnel junctions, the importance of an asymmetric dispersion was pointed out (Hatta, Nagao and

Mukaa 1996); they used modified two-band model, in which the different conduction- and valence-band edge effective masses are taken into account. In Figs. 4–6 the complex bands for SiO_2 do not show strong asymmetry despite the large difference between the conduction- and valence-band edge effective masses. However, the asymmetry would change the currents exponentially, and hence we expect that the asymmetric two-band model improves our tunneling calculation; there exists some discrepancies between calculated and measured results in Fig. 3. Furthermore, in the aggressively scaled MOSFETs the carriers tunnel through the wide range of the SiO_2 gap region, and DT of hot carriers as well as cold carriers are important to analyze the device reliability (Deguchi *et al.* 2000). In particular, the hot holes pass through the bottom of the gap region, where the complex band structure is very complicated. For more accurate DT simulation applicable to a variety of situations, it is important to take account the realistic complex band structures.

5. Conclusion

In summary, we have presented the simulation of the tunneling current through the ultra-thin gate oxides (2.6–3.4 nm). By using the two-band model for the complex band structure of SiO_2 , good agreements were obtained between calculated and experimental tunneling currents measured by the charge separation technique. It has been also demonstrated that the two-band model reflects the essential characteristics of the realistic complex band structure obtained from the tight-binding calculation.

Acknowledgments

This work was supported by the Semiconductor Technology Academic Research Center (STARC).

References

- Boykin T.B. 1996. Phys. Rev. B 54: 7670.
- Deguchi K., Ishida A., Uno S., Kamakura Y., and Taniguchi K. 2000. Appl. Phys. Lett. 77: 1384.
- Ghetti A., Hamad A., Silverman P.J., Vaidya H., and Zhao N. 1999. In: Proceedings of 1999 International Conference on Simulation of Semiconductor Processes and Devices, Kyoto, p. 239.
- Gnani E., Reggiani S., Colle R., and Rudan M. 2000. IEEE Trans. Electron Devices 47: 1795.
- Hatta E., Nagao J., and Mukaa K. 1996. J. Appl. Phys. 79: 1511.
- Kane E.O. 1966. Semiconductors and Semimetals 1: 75.

- LaFemina J.P. and Duke C.B. 1991. *J. Vac. Sci. Technol. A* 9: 847.
- Lee W.C. and Hu C. 2001. *IEEE Trans. Electron Devices* 48: 1366.
- Rana F., Tiwari S., and Buchanan D.A. 1996. *Appl. Phys. Lett.* 69: 1104.
- Schulman J.N. and Chang Y.C. 1983. *Phys. Rev. B* 27: 2346.
- Städele M., Tuttle B.R., and Hess K. 2001. *J. Appl. Phys.* 89: 348.
- Takagi S., Takayanagi M., and Toriumi A. 1999. *IEEE Trans. Electron Devices* 46: 1446.
- Ting D.Z.-Y., Yu E.T., and McGill T.C. 1992. *Phys. Rev. B* 45: 3583.
- Tsu R. and Esaki L. 1974. *Appl. Phys. Lett.* 22: 562.
- Vögl P., Hjalmarson H.P., and Dow J.D. 1983. *J. Phys. Chem. Solids* 44: 365.



Comparison of Quantum Corrections for Monte Carlo Simulation

BRIAN WINSTEAD

Beckman Institute and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

HIDEAKI TSUCHIYA

Department of Electrical and Electronics Engineering, Kobe University 1-1, Rokko-dai, Nada-ku, Kobe 657-8501, Japan

UMBERTO RAVAIOLI

Beckman Institute and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Abstract. As semiconductor devices are scaled down to nanometer scale dimensions, quantum mechanical effects can become important. For many device simulations at normal temperatures, an efficient quantum correction approach within a semi-classical framework is expected to be a practical way applicable to multi-dimensional simulation of ultrasmall integrated devices. In this paper, we present a comparative study on the three quantum correction methods proposed to operate within the Monte Carlo framework, which are based on Wigner transport equation, path integrals, and Schrödinger equation. Quantitative comparisons for the strengths and weaknesses of these methods are discussed by applying them to size quantization and tunneling effects.

Keywords: ultrasmall MOSFET, Monte Carlo methods, nanotechnology, quantum correction, Schrödinger equation, quantum effects

1. Introduction

As semiconductor devices are scaled down to nanometer scale dimensions, quantum mechanical effects can become significant, and a full quantum transport model is necessary if coherent effects dominate device behavior. However, for many practical devices, an efficient alternative is to include quantum corrections within a semi-classical framework. If a physically-based model such as Monte Carlo is used, it is easier to include important transport physics than in most available quantum transport approaches. For example, in MOSFETs scaled below 100 nm, bandstructure and scattering mechanisms must still be modeled to a certain degree of sophistication, while coherence effects should only play a secondary role because the potential profiles along the transport path are typically smooth,

minimizing quantum mechanical reflections. Instead of coherent transport, the major quantum effects to be concerned about in this case are size quantization and tunneling. Size quantization can be captured with quantum corrections because in the direction perpendicular to the transport, the device is essentially in quasi-equilibrium conditions, and the major issue is to adjust the statistical occupation probabilities. Tunneling occurs in the direction of transport, but for sufficiently wide or high single barriers, the quantum region of action can be assumed to be strongly localized in the neighborhood of the barrier itself.

Quantum corrections can be incorporated into a semi-classical Monte Carlo simulator by introducing a quantum potential term which is superimposed onto the classical electrostatic potential seen by the simulated particles. The essence of the technique is illustrated

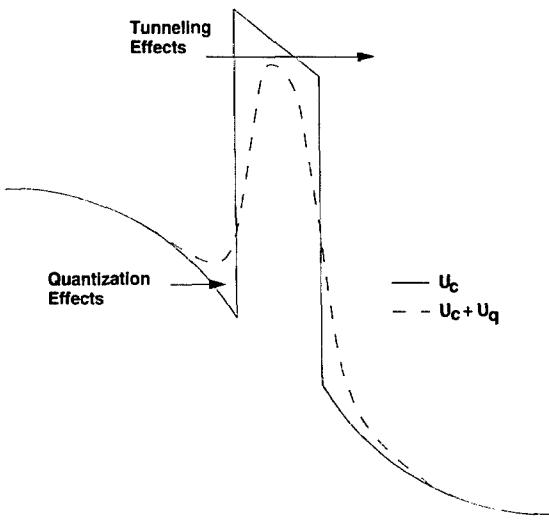


Figure 1. Illustration of how quantum effects are treated by adding a “quantum potential” to the electrostatic potential.

pictorially for a single tunneling barrier in Fig. 1. Raising a particle’s potential energy in a quantum well, or lowering it at the top of a barrier can modify the semi-classical transport, thus reproducing to first-order the average effects of quantization and tunneling on the carrier distribution.

Several quantum correction approaches are possible. These procedures in general entail the self-consistent calculation of a correction potential which is added to the semi-classical solution. Approximate quantum models are used to obtain the corrected potential from the semi-classical potential itself, to steer the transport toward a situation that mimics as much as possible the quantum behavior. The methods proposed to operate within the Monte Carlo framework include methods based on Wigner equation (Tsuchiya and Miyoshi 1999), path integrals (Ferry 2000), and Schrödinger equation (Winstead and Ravaioli 2001). The goal of this paper is to review comparatively these three main approaches, underscoring the strengths and weaknesses of each of them. Quantitative comparisons are presented to help in understanding for which applications one method might be more efficient or appropriate over the others.

2. Description of Quantum Corrections

2.1. Wigner-Based Correction

The Wigner-based quantum correction can be derived starting from a suitable form of the Wigner transport

equation (Wigner 1932)

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f - \frac{1}{\hbar} \nabla_r V \cdot \nabla_k f + \sum_{\alpha=1}^{\infty} \frac{(-1)^{\alpha+1}}{\hbar 4^\alpha (2\alpha + 1)!} (\nabla_r \cdot \nabla_k)^{2\alpha+1} V f = \left(\frac{\partial f}{\partial t} \right)_c \quad (1)$$

Here, k is the crystal momentum, V is the classical potential and the term on the right-hand side represents the effects of collisions. The non-local quantum mechanical effects are represented in the fourth term on the left-hand side of (1). In the limit of slow spatial variations, the non-local terms disappear and we recover the standard Boltzmann Transport equation (BTE). The simplest approach to quantum correction is to start by using only the lowest order term with $\alpha = 1$ in the summation. Following this approximation, one obtains an equation that closely resembles the structure of the BTE, with one additional term providing a quantum correction. This quantum corrected BTE takes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f - \frac{1}{\hbar} \nabla_r V_w \cdot \nabla_k f = \left(\frac{\partial f}{\partial t} \right)_c \quad (2)$$

where the term V_w contains the quantum potential. V_w depends on the distribution function, which in turn can be resolved numerically by Monte Carlo simulation, for equilibrium or non-equilibrium cases. We take here a simpler approach, which assumes a drifted Maxwellian distribution function with parabolic dispersion relation. It allows us to represent V_w with an analytical form. Limiting the derivation to one dimension for clarity, V_w becomes (Tsuchiya and Ravaioli 2001)

$$V_w(k, n) = V + \frac{k_B T}{24} [\gamma^2 (k - \bar{k})^2 - 3\gamma] \frac{\partial^2 \ln(n)}{\partial x^2},$$

$$\gamma = \frac{\hbar^2}{m^* k_B T} \quad (3)$$

where n is the carrier concentration. In (3), the corrected potential, V_w , depends on both the location and the momentum of the individual particles. A simplified version of V_w can also be derived by assuming in (3) a thermal equilibrium average energy as (Tsuchiya and Miyoshi 1999, Tsuchiya and Ravaioli 2001)

$$V_w(n) = V - \frac{\hbar^2}{12m^*} \frac{\partial^2 \ln(n)}{\partial x^2} \quad (4)$$

This simplified, momentum-independent formulation has some advantage over the more complex

momentum-dependent version (3) because in addition to use in Monte Carlo, it can be applied for quantum corrections in lower levels of the transport simulation hierarchy such as hydrodynamic (Zhou and Ferry 1993) or drift-diffusion (Ancona and Iafrate 1989). We have to add that for a multi-dimensional problem the Wigner-based correction should be represented in terms of a quantum force correction, not a quantum potential correction (Tsuchiya and Ravaoli 2001).

2.2. Effective Potential Correction

The effective potential approach to quantum correction was developed by Feynman and Hibbs (1965). To derive the effective potential, a variational method can be used to calculate contribution to the path integral of a particle's quantum fluctuations around its classical path. Using a trial potential to first order in the average point on each path, the effective classical potential becomes

$$V_{eff}(x) = \frac{1}{\sqrt{2\pi}a} \int_{-\infty}^{\infty} V(x_0) e^{-\frac{(x_0-x)^2}{2a^2}} dx_0, \quad (5)$$

$$a^2 = \frac{\hbar^2}{12m^*k_B T}$$

Equation (5) represents a smearing of the electrostatic potential on a length scale of the parameter, a , which can also be interpreted as the effective quantum mechanical "size" of the particle (Ferry 2000).

Feynman later improved this simple correction using a second-order trial potential (Feynman and Kleinert 1986) which yields the following effective potential, W_1

$$W_1(x_0) = \min_{a^2(x_0), \Omega(x_0)} \{ \tilde{W}_1(x_0, a^2(x_0), \Omega(x_0)) \}, \quad (6)$$

$$\text{where } V_{a^2}(x) \equiv \frac{1}{\sqrt{2\pi}a} \int_{-\infty}^{\infty} V(x_0) e^{-\frac{(x_0-x)^2}{2a^2}} dx_0$$

$$\tilde{W}_1(x_0, a^2, \Omega) = \frac{1}{\beta} \ln \frac{\sinh(\beta\Omega/2)}{\beta\Omega/2} - \frac{\Omega^2}{2} a^2 + V_{a^2}(x_0)$$

V_{eff} in (6) corresponds to W_1 in (6) with the special non-optimal choice of $\Omega \equiv 0$. A typical solution of W_1 for MOS quantization effects is indicated in Fig. 2. For this application, the benefits of using the W_1 effective potential relative to the simpler V_{eff} with a allowed to vary as a tuning parameter appear to be marginal. For

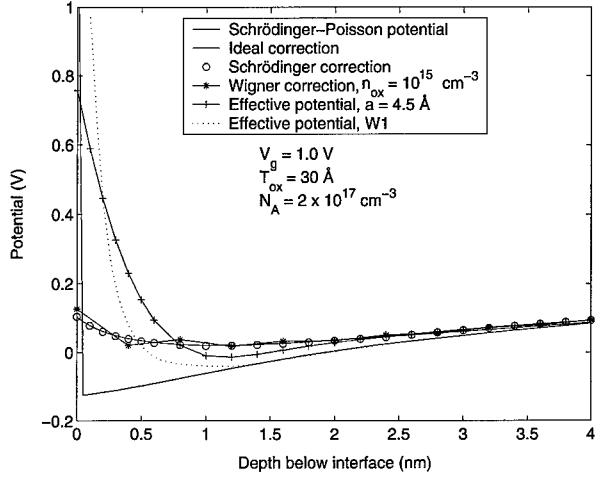


Figure 2. Typical behavior of the quantum potential for an MOS capacitor using several different quantum correction approaches.

practical application, the primary focus in this work will be on the V_{eff} version of the correction.

2.3. Schrödinger-Based Correction

In the Schrödinger-based approach for quantum correction, the Schrödinger equation is solved periodically in a simulation using the self-consistent electrostatic potential as input. In contrast to the Wigner-based and effective potential corrections, the quantum potential in this method is calculated from the exact energy levels and wavefunctions corresponding to the electrostatic potential solution. The first step in the procedure is to calculate the overall shape of the quantum density by filling the energy levels according to an equilibrium Maxwell-Boltzmann distribution. This quantum density shape is mapped to a quantum potential through

$$V_{schr}(z) = -kT \ln(n_q(z)) - V_p(z) + V_0 \quad (7)$$

Here, V_{schr} is the quantum correction, z is the direction normal to the interface, n_q is quantum density from the Schrödinger equation or equivalently the converged Monte Carlo concentration, V_p is the potential from the Poisson solution, and V_0 is an arbitrary reference potential determined by the knowledge that the correction should go to zero away from the quantum region, where the behavior is semi-classical. Only the shape of the quantum density is used, therefore, one does not need to invoke the exact Fermi level in the calculation. In this way the correction can be adapted to

treat nonequilibrium device simulation (Winstead and Ravaioli 2001).

The quantum-corrected potentials, V_w , V_{eff} , or V_{schr} , differ in their method of calculation and their underlying assumptions. However, they are all incorporated into a Monte Carlo simulation in a similar way. As a Monte Carlo simulation evolves in time, the corrections are recalculated along with the Poisson equation to maintain self-consistency. The quantum-corrected potential is then used instead of the electrostatic potential to calculate the forces on the Monte Carlo particles. Other than this modification of the classical forces applied to the particles, the quantum-corrected Monte Carlo simulation can be carried out in the same manner as the uncorrected case.

3. Quantization Effects

To study quantization effects, the models described in the preceding section were implemented in the 2-D full-band Monte Carlo simulator, MOCA (Duncan, Ravaioli and Jakumeit 1998). Because of its technological importance as a building block for devices, the MOS capacitor was used as a prototype structure for this comparative study. For verification, the quantum mechanical charge density and potential were also calculated using self-consistent Schrödinger/Poisson simulation.

Figure 2 illustrates the typical behavior of the quantum potential for the different methods. Here, the "ideal" quantum potential is the correction which would exactly reproduce the quantum density from the Schrödinger-Poisson solution. The results indicate that the Schrödinger-based correction provides the most accurate model, which closely matches the ideal value with no fitting parameters. This is expected because the approaches makes use of a complete solution of the Schrödinger equation instead of an approximate quantum solution. In addition, since there are no fitting parameters, the accuracy of the method is not sensitive to variations in the physical parameters of the MOS capacitor. Figures 3 and 4 compare the detailed solutions for concentration obtained from a full quantum calculation and from a Schrödinger-corrected Monte Carlo simulation, over a wide range of gate biases and for substrate dopings of $N_A = 2 \times 10^{17} \text{ cm}^{-3}$ and $N_D = 2 \times 10^{17} \text{ cm}^{-3}$.

The Wigner-based quantum potential is also found to be accurate for quantization effects in the MOS capacitor, if a fitting parameter is used for the density at

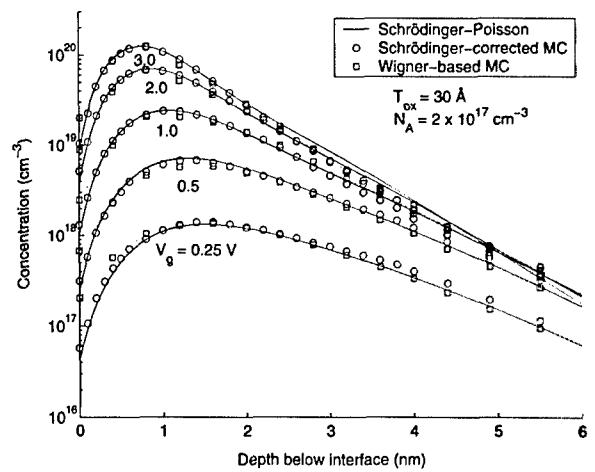


Figure 3. Electron concentration distributions in an inverted MOS capacitor from two different quantum-corrected Monte Carlo and self-consistent Schrödinger-Poisson methods over a range of gate bias.

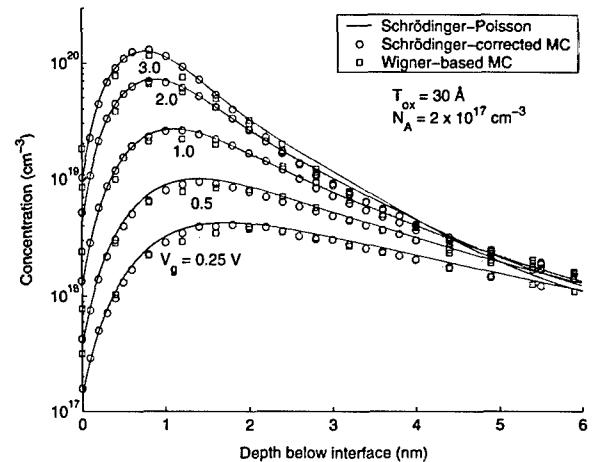


Figure 4. Electron concentration distributions in an accumulated MOS capacitor from two different quantum-corrected Monte Carlo and self-consistent Schrödinger-Poisson methods over a range of gate bias.

the interface. Results obtained using the Wigner correction method are also shown in Figs. 3 and 4. The fitting parameter used here is an empirical charge layer of $1 \times 10^{15} \text{ cm}^{-3}$ which is included in the oxide region for the calculation of the correction at the interface point. Beside this necessary adjustment at the interface, the quantum correction (4) is applied with no additional fitting parameters. This scheme allows for the proper adjustment of the interface density for a wide range of biases and doping while giving a reasonably accurate quantum density elsewhere.

For the Feynman effective potential given by (6), the "size" parameter, a was treated as an empirical fitting parameter, as suggested by Ferry (2000). The best fit value for the size parameter in the MOS structures studied here was found to be $a = 4.5 \text{ \AA}$. The effective potential method is accurate in reproducing integrated quantities. Figure 5 shows the total sheet charge for a Monte Carlo simulation of the MOS capacitor with the effective potential correction, and Fig. 6 shows, for the same simulation, the average displacement of the carriers from the Si/SiO₂ interface, which is indicative

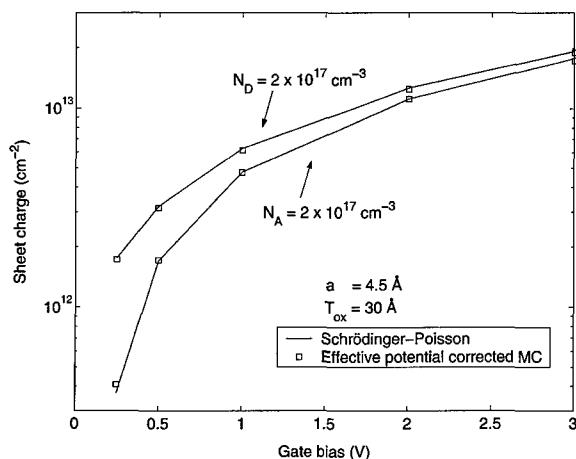


Figure 5. Sheet charge density in a MOS capacitor as a function of gate bias calculated with effective potential corrected Monte Carlo and self-consistent Schrödinger-Poisson methods.

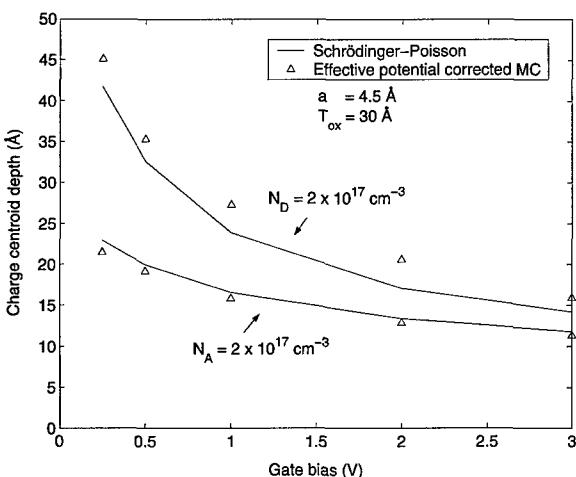


Figure 6. Location of charge centroid in a MOS capacitor as a function of gate bias calculated with effective potential corrected Monte Carlo and self-consistent Schrödinger-Poisson methods.

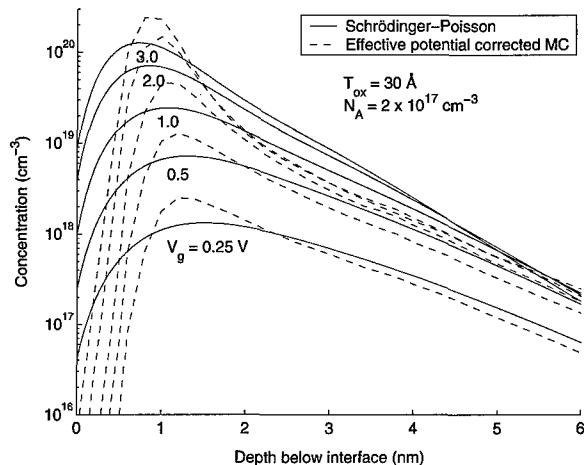


Figure 7. Electron concentration distributions in an inverted MOS capacitor calculated with effective potential Monte Carlo and self-consistent Schrödinger-Poisson methods.

of the quantum repulsion. However, if the detailed spatial behavior of the effective potential correction is analyzed, one can see significant deviations from the quantum solution. Figure 7 shows the detailed concentration under the gate of the MOS capacitor. Typically, the correction is very large at the interface, leading to a layer of width $\sim a$ next to the oxide interface, where the concentration is significantly lower than that is expected by the quantum solution. Compensating for that, the correction becomes smaller than the expected one at the deeper location inside the substrate, leading to typically a larger peak concentration than the quantum solution.

It can be shown theoretically that the momentum-independent Wigner-based method (4) is an approximation to the effective potential (Ferry 2000). However, the simulation results presented here indicate that the momentum-independent Wigner-based correction gives a solution which is substantially closer to the detailed quantum behavior. This is due to the fact that the Wigner correction is local, while the effective potential correction is non-local. Neither correction is strictly valid at a heterojunction. However, a single parameter can be used to fit the singularity at the interface for the Wigner correction, since it is local. The silicon region in which the transport actually occurs has a more slowly-varying potential than in the neighborhood of the interface, and thus no fitting is necessary. The application of a non-local effective potential act differently in the overall correction schemes. The adjustment of a fitting parameter to accommodate the strong influence

of the interface on the overall solution requires a compensation in the silicon region where the solution has to deviate to maintain the averages.

In addition to accuracy, another important consideration in practical Monte Carlo device simulation is the execution time. For all three methods, the CPU time required to calculate the corrections is negligible relative to the overall Monte Carlo simulation time. However, there is an important difference, in the fact that the Schrödinger-based correction and the effective potential correction are calculated using the electrostatic potential as input, while the Wigner-based correction is calculated from concentration. The noise in the Monte Carlo concentration estimator is always higher than for the potential, and a Wigner-corrected Monte Carlo simulation can take significantly longer time to converge than an uncorrected semi-classical Monte Carlo simulation. In contrast, adding the Schrödinger-based or the effective potential correction to the Monte Carlo procedure does not increase total cpu time in a very significant way. A self-consistent simulation with the full-band MOCA code using 30000 particles and a non-uniform grid of 300×200 nodes for the Poisson equation requires approximately 80 MB of RAM. On a standard 800 Mhz Intel processor, approximately 1000 iterations per hour are executed, where one iteration normally corresponds to a time step of 1 fs or less.

4. Tunneling Effects

For the purpose of studying quantum corrections in the context of tunneling, the Wigner-based correction and the effective potential were implemented into a 1D GaAs/AlGaAs Monte Carlo simulator. For this case, the Schrödinger correction was not applied, since it is best suited for capturing quantum confinement effects. The tunneling test structure consists of a 4-nm wide GaAs/AlGaAs single barrier with a conduction band discontinuity of 0.22 eV and a temperature of 300 K.

As shown previously, the effective potential correction encounters difficulties in the neighborhood of the abrupt transition between oxide and silicon in the MOS system, since there is very large energy jump of about 3.1 eV and the underlying assumptions behind the theory tend to break down. The problem should not be as severe in the presence of smaller barriers, as is the case for the GaAs/AlGaAs system and the effective potential should be a very good candidate for practical inclusion of tunneling effects.

In applying the Wigner-based correction to MOS quantization, the difficulties near the large barrier were overcome by tuning the correction at the interface point. However, for tunneling it is necessary to model transport on both sides of the interface, and this scheme, that is based on assuming a concentration layer, cannot be used. Instead, for tunneling simulations we apply the theoretical value of the Wigner correction. In order to increase the accuracy, here we implement the momentum-dependent method (3) in addition to the momentum-independent method (4) used in the quantization simulations.

For the tunneling simulation the bias was varied from 0 to 0.3 V, and the GaAs effective mass of $m^* = 0.067m_0$ was used in all three corrections. From (6), this corresponds to a value of 1.9 nm for a in the effective potential. To benchmark the results, the quantum tunneling current was also calculated using a transfer matrix method (Brennan and Summers 1987). Figure 8 plots the resulting current from the transfer matrix and Monte Carlo methods. All of the quantum corrected results improve significantly upon the classical simulation. The momentum-dependent Wigner correction and the effective potential are the more accurate methods. However the details of their results differ, which is expected because each method stems from a different set of assumptions. The momentum-independent Wigner method is less accurate, which is consistent with the fact that it can be considered an approximation to either the momentum-dependent method or to the effective potential. These same trends

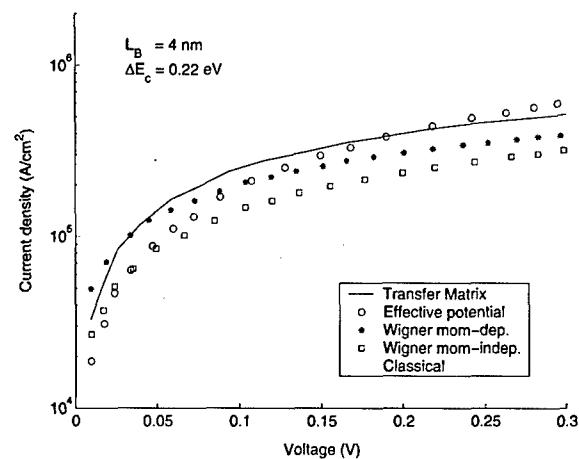


Figure 8. Tunneling current in a 4-nm GaAs-AlGaAs tunneling barrier over a range of bias calculated with three quantum-corrected Monte Carlo and transfer matrix methods.

also hold for other small tunneling barriers, such as possible source-drain tunneling in highly scaled MOSFETs. If an improved accuracy is desired, the α parameter in the effective potential, or equivalently, the m^* parameter in the Wigner-based corrections can be adjusted for a best fit. However, in such a case, recalibration of the fitting parameters may be required when different barriers are considered.

5. Conclusions

Three methods for introducing quantum corrections in semi-classical Monte Carlo simulation have been studied and compared. For the size-quantization case in the MOS system, the Schrödinger-based correction has some intrinsic advantage, since this method does not require fitting parameters, it is accurate, and it adds only negligible CPU time to a Monte Carlo simulation. In contrast, while the Wigner-based method can be tuned to be as accurate, it is in general slower and it requires a fitting parameter. The effective potential method reproduces reasonably well integrated quantities related to size quantization, but it is spatially inaccurate even if the fitting parameter is optimized. For the tunneling case, the Schrödinger-based correction is not appropriate. Instead, the momentum-dependent Wigner correction or effective potential methods can be used with similar accuracy. In this case, the effective potential should have a computational advantage, since the Wigner correction is more affected by noise. On the other hand, the Wigner formulation can still be useful for detailed physical investigations, since it can be extended to include momentum-dependent distributions. One could introduce an analytical distribution function or even a numerical one evaluated directly with the semi-classical Monte Carlo procedure.

Acknowledgment

This work was partially supported by the Semiconductor Research Corporation, contract 99-NJ-726. H. T also thanks for the support by the Ministry of Education, Science, Sports and Culture of Japan, Grand-in-Aid for Encouragement of Young Scientists, 13750061, 2001.

References

- Ancona M. and Iafrate H. 1989. Quantum correction to the equation of state of an electron gas in a semiconductor. *Phys. Rev. B* 39: 9536–9539.
- Brennan K. and Summers C. 1987. Theory of resonant tunneling in a variably spaced multiquantum well structure: An Airy function approach. *J. Appl. Phys.* 61: 614–623.
- Duncan A., Ravaoli U., and Jakumeit J. 1998. Full-band Monte Carlo investigation of hot carrier trends in the scaling of metal-oxide-semiconductor field-effect transistors. *IEEE Trans. Electron Devices* 45: 867–876.
- Ferry D.K. 2000. Effective potentials and the onset of quantization in ultrasmall MOSFETs. *Superlattices and Microstructures* 28(5/6): 419–423.
- Feynman R.P. and Hibbs A.R. 1965. *Quantum Mechanics and Path Integrals*. McGraw-Hill, New York.
- Feynman R. and Kleinert H. 1986. Effective classical partition functions. *Phys. Rev. A* 34(6): 5080–5084.
- Tsuchiya H. and Miyoshi T. 1999. Quantum transport modeling of ultrasmall semiconductor devices. *IEICE Trans. Electron.* E82-C(6): 880–887.
- Tsuchiya H. and Ravaoli U. 2001. Particle Monte Carlo simulation of quantum phenomena in semiconductor nanostructures. *J. Appl. Phys.* 89: 4023–4029.
- Wigner E. 1932. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* 40: 749–759.
- Winstead B. and Ravaoli U. 2001. A coupled Schrödinger/Monte Carlo technique for quantum-corrected device simulation. In: *Device Research Conference*, Notre Dame, Indiana.
- Zhou J.R. and Ferry D.K. 1993. Modeling of quantum effects in ultrasmall HEMT devices. *IEEE Trans. Electron Devices*. 40: 421–427.



Monte Carlo Based Calculation of the Electron Dynamics in a Two-Dimensional GaN/AlGaN Heterostructure in the Presence of Strain Polarization Fields

TSUNG-HSING YU AND KEVIN F. BRENNAN

School of Electrical and Computer Engineering, Georgia Tech, Atlanta, GA 30332-0250, USA

Abstract. We report on the workings of a Monte Carlo based simulator useful for studying electron transport in two-dimensional systems. The simulator utilizes a self-consistent solution of the Schroedinger and Poisson equations to obtain the allowed two-dimensional energy levels, band bending and electronic wavefunctions. Defect scattering through interface roughness and ionized impurities along with lattice scattering arising from polar optical, deformation potential and piezoelectric interactions are included in the model. The two-dimensional scattering rates are calculated using the numerically determined wavefunctions. The final state following polar optical scattering is determined numerically based on the two-dimensional physics of the process. The model further includes the effects of strain induced polarization fields present in GaN/AlGaN heterostructures. Transfer into the AlGaN layer and its effects are also considered. Calculations are presented for the steady-state velocity showing the importance of the new two-dimensional final state selection technique. Additionally, calculations are presented that show the importance of the strain polarization fields.

Keywords: heterostructures, two-dimensional transport, Monte Carlo

1. Introduction

Transport within a two-dimensional system is well known to be substantially different from that in a three-dimensional system (Yokoyama and Hess 1986, Kawamura and Das Sarma 1992). The electronic structure along with the scattering rates are quite different between two and three dimensional systems (Ridley 1997) and as a result the electron dynamics are different. In strained semiconductor systems, such as $\text{GaN}_{\text{Al}_{1-x}\text{Ga}_x\text{N}}$, the strain produces polarization fields that alter the band bending, electronic structure and scattering rates. Though there has been some attempt to model this effect in the III-nitrides, most of the studies have been made only for the zero field mobility (Hsu and Walukiewicz 2001a, Yu and Brennan 2001b). The field-dependent drift velocity and concomitant mobility are of greater interest in simulating HFET devices. To the authors' knowledge, there has been only one work that has examined the field dependence of the drift velocity

and mobility in AlGaN/GaN heterojunctions (Li, Joshi and Fazi 2000). However, in the work by Li, Joshi and Fazi (2000) the electronic structure in the two-dimensional system was approximated using a simple triangular well approximation, the scattering rates were calculated using approximate forms for the wave functions and only two subbands were included in the analysis. In the present work, we describe a fully numerical approach in which the subbands and corresponding scattering rates are determined self-consistently from the solution of the Poisson and Schroedinger equations. Our model includes multiple subbands with the number depending upon the band bending, both spontaneous and piezoelectrically induced polarization effects, and the possibility of real space transfer into the AlGaN. Additionally, we present a technique for determining the final state following a two-dimensional polar optical phonon scattering event.

It is the purpose of this paper to outline the computational details of the workings of our fully numerical

two-dimensional simulator. During the course of this work we have developed a new numerical technique for the final state selection following two-dimensional polar optical phonon scattering.

2. Model Description

The two-dimensional transport is solved using the ensemble Monte Carlo technique with a parabolic band approximation. The subband energies and electronic wavefunctions are determined numerically from a self-consistent solution of the Schrödinger-Poisson equation. The scattering rates are then determined numerically by computing the appropriate matrix elements using the numerical wavefunctions. The scattering mechanisms included in the analysis are polar optical and acoustic phonon, piezoelectric, remote ionized impurity and interface roughness scattering. The calculations presented here are all made at 300 K where polar optical phonon scattering dominates. Therefore, it is highly important to properly treat the effects of polar optical scattering in the simulation.

During the course of the Monte Carlo simulation the final state selection following polar optical scattering is determined using a new technique that fully embraces the physics of two-dimensional transport. The polar optical phonon scattering angle in two-dimensional electron gases can be calculated by the formula derived from the summation probability of scattering between 0 and θ , assuming θ is the same for positive and negative,

$$\gamma = \frac{\int_0^\theta \frac{H_{mn}(q) dq}{q}}{\int_0^\pi \frac{H_{mn}(q) dq}{q}} \quad (1)$$

where $H_{mn}(q)$ is the subband coupling coefficient and q is the phonon wavevector component parallel to the layer plane. These quantities are defined in Yokoyama and Hess 1986. For a given angle θ we can calculate the corresponding probability γ between 0 and 1 from the above formula. To implement this scheme within the Monte Carlo simulation, we have to reverse the calculation sequences. Therefore, we utilize a piecewise linear approximation to fit the curve and hence with a random number γ we can find its corresponding scattering angle θ . Since the polar optical phonon scattering is strongly dependent on the electron energy, in principle we should calculate the random number vs. polar angle for each energy value to obtain the precise polar angles. However, the computational demands of such an approach are presently

overwhelming. Alternatively, we calculate the random number vs. polar angle for ten energy values to build a look-up table. For each value of the carrier energy, the polar angle is estimated using the curve with the nearest value of the carrier energy.

The importance of this new technique is illustrated by Fig. 1. Figure 1 shows a comparison between the two-dimensional (2D) and three-dimensional (3D) bulk polar optical phonon emission and absorption scattering angles calculated for a carrier with a total energy of 400 meV relative to the conduction band minimum. In the 3D model, the final state of the carrier is found using the analytical model provided by Tomizawa (1993). For a given random number, the 2D emission angle is larger than the corresponding angle calculated using the 3D approximation. As a result, a significant difference in the calculated final state and consequently carrier velocity can occur between the 2D and 3D models as will be shown below. Therefore, it is important to incorporate the correct 2D polar angle calculation to accurately determine the 2D final state selection in the Monte Carlo simulation. Usage of the 3D formulation to find the final state after a 2D polar optical scattering is thus inadequate.

Given that the scattering rates and energy levels are distinctly different between the 2D and 3D systems, it is necessary to define when the carrier is in either system. Physically, as the energy increases within the 2D system the subbands become increasingly closer together ultimately producing a quasi-continuum 3D system. In our calculation, we determine sufficient energy bands until the energy separation between successive higher energy subbands in the quantum well is less than the thermal energy. The maximum energy subband is then defined as the threshold energy separating the 2D and 3D systems. A carrier with energy above the threshold is treated as belonging to the 3D system. The transition between the 2D system and the bulk GaN and AlGaN is accomplished through both carrier drift and polar optical phonon scattering (Brennan and Park 1989, Park and Brennan 1989). Once the energy of the carrier within the 2D system approaches the threshold energy, the electron can acquire sufficient energy from the applied electric field or via the polar optical phonon absorption to enter the 3D states. Similarly, the electron can transfer to the 2D system from the bulk by polar optical phonon emission or drifting downwards during the drift motion.

Once the electrons are heated up to the three-dimensional states in GaN, they can acquire sufficient

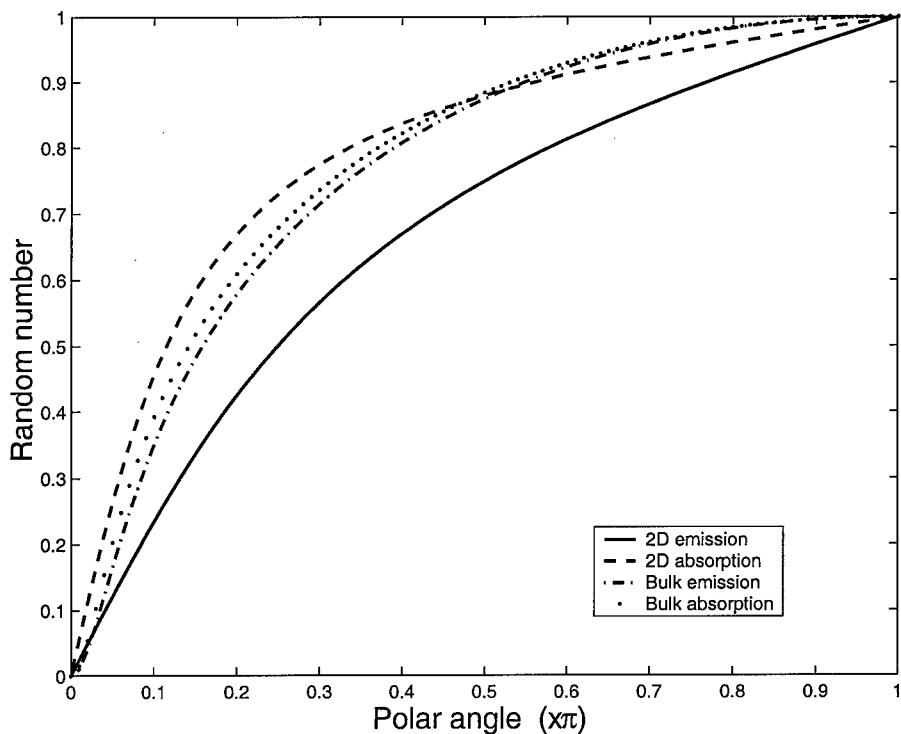


Figure 1. Calculated 2D polar optical phonon scattering angle in the $\text{Al}_{0.2}\text{Ga}_{0.8}\text{N}/\text{GaN}$ structure is compared to the scattering angle in bulk GaN. The calculated carrier energy is 400 meV relative to the conduction band minimum.

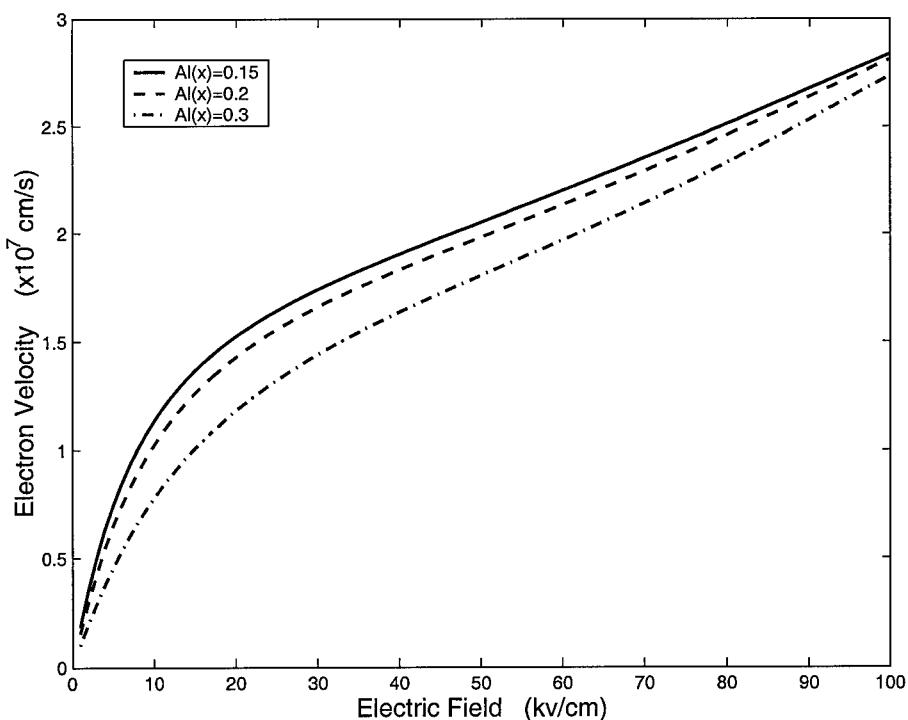


Figure 2. Calculated electron velocity vs electric field for AlGaN/GaN HFET structures with Al mole fractions of 15%, 20%, and 30%.

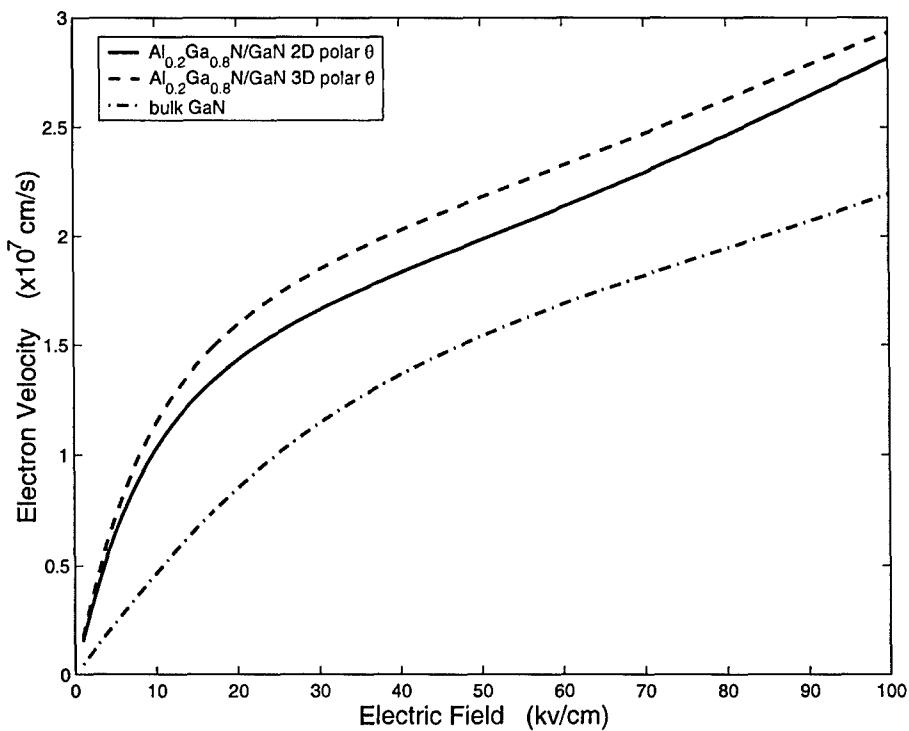


Figure 3. Comparison between 2D and bulk calculated electron velocity. Two different final state selections, 2D and bulk polar angles, are calculated in an $\text{Al}_{0.2}\text{Ga}_{0.8}\text{N}/\text{GaN}$ HFET structure.

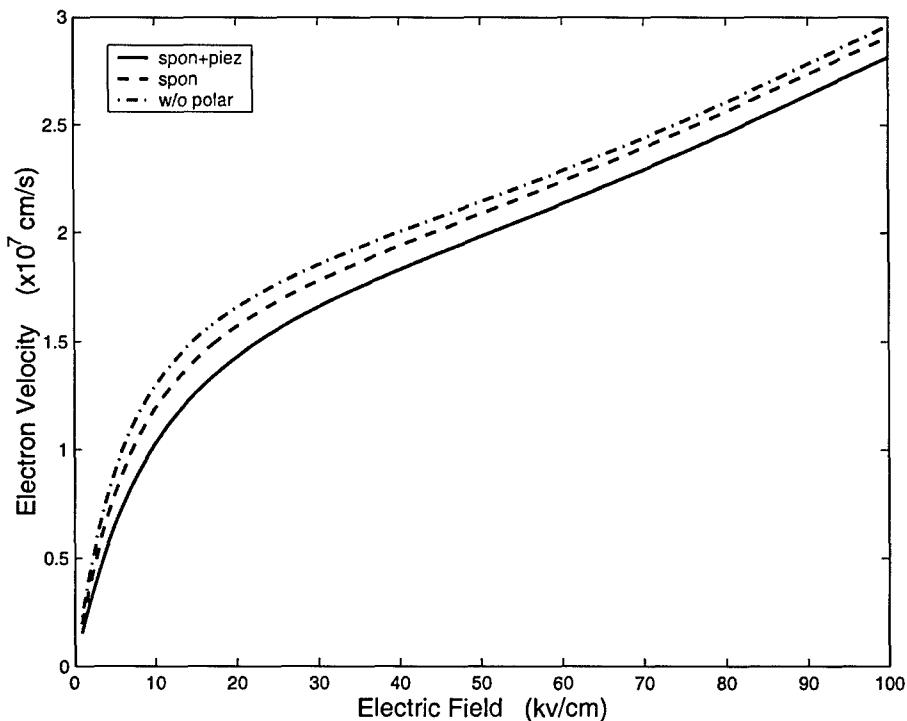


Figure 4. Calculated electron velocity as a function of electric field in an $\text{Al}_{0.2}\text{Ga}_{0.8}\text{N}/\text{GaN}$ HFET structure with and without polarization effects.

energy from the applied electric field along the heterojunction to populate higher valleys or transfer to the adjacent AlGaN layer. Real space transfer between these two materials occurs only when the incident and transmitted wavevectors of the carrier satisfy the phase matching condition (Gaylord and Brennan 1989) at the heterointerface, otherwise the carriers will be reflected. Moreover, the transverse field due to the band bending in the heterostructure is included in our calculation and the carrier drift motion along the transverse direction is required to satisfy energy and momentum conservation.

3. Simulation Results

The device structure and material parameters for bulk GaN used in the simulation are listed in Table II of Yu and Brennan (2001b) and Table I of Farahmand *et al.* (2001) respectively. The effective masses and intervalley energy separation for the $\text{Al}_x\text{Ga}_{1-x}\text{N}$ ternary compounds are extracted from a pseudopotential band structure calculation (Goano *et al.* 2000a, 2000b) by using a linear interpolation. The interface roughness parameters are adopted to be the same as in Yu and Brennan (2001b) to maintain a best fit to the experimental data for the zero-field mobility.

Figure 2 shows the steady state electron velocity versus the applied electric field along the channel direction for different Al mole fractions. To check the validity of our model, we have compared our theoretical calculations to the simulation results given in Li, Joshi and Fazi (2000). For the 15% Al composition at room temperature, our simulation results are in good qualitative agreement with Li's calculation (Li, Joshi and Fazi 2000) over the electric field range from 10 to 50 kV/cm. As shown in an earlier calculation (Yu and Brennan 2001b), the Al compositions can be used to control the magnitude of polarization field in the heterointerface. Owing to an increase of the Al mole fraction, the magnitude of the polarization field and the conduction band offset will increase simultaneously. The former will induce a significantly larger sheet charge density and the latter will enhance the confinement of carriers in the well. Therefore, the strong band bending caused by the larger sheet charge density will push the 2D electrons closer to the heterojunction interface. As a result, the effects of interface roughness scattering are enhanced leading to degradation in the electron velocity as the Al composition increases as can be seen from inspection of Fig. 2.

Comparison is made between the 2D and 3D polar optical phonon scattering angle models for the electron velocity in the $\text{Al}_{0.2}\text{Ga}_{0.8}\text{N}/\text{GaN}$ heterostructures. The calculated results for this comparison are shown in Fig. 3. Since the 2D polar angle is quite sensitive to the electron energy, a carrier with a higher energy will be scattered by a smaller angle and vice versa. On the other hand, the polar angles calculated using the 3D formulation stay within a narrow range for different electron energy. For example, as the electron total energy increases from 0.4 eV to 0.8 eV relative to the conduction band minimum, the 2D emission polar angle for the first intrasubband scattering decreases from 0.25π to 0.05π . The deviation of 2D angles is 0.2π compared to 0.05π in 3D case. Because the 3D polar angles on average are less than the 2D angles, the electron velocity is overestimated by using 3D final state selection for polar optical phonon scattering as shown in Fig. 3. As can be seen from the figure there is a significant difference in the calculated velocity between the two models. In addition, we plot the electron velocity for bulk GaN in Fig. 3. Due to the spatial separation between the doped donors and the free carriers in modulation doped heterostructures, the 2D electron velocity is significantly larger than that of the bulk as expected.

The influence of polarization effects on the electron velocity in an $\text{Al}_{0.2}\text{Ga}_{0.8}\text{N}/\text{GaN}$ heterostructure is shown in Fig. 4. Three cases are examined in Fig. 4. These are without polarization, spontaneous polarization only, and the strained case including both spontaneous and piezoelectric polarization. The polarization field in the strained case is larger than that for the other two cases. Inspection of Fig. 4 shows that as the polarization field increases the electron drift velocity decreases. The polarization field induces a larger sheet charge density in the quantum well and also causes a strong band bending in the conduction band profile. Therefore the transverse electric field due to the conduction band bending will dramatically increase. As a result, the electrons in the channel are pushed closer to the heterointerface resulting in an increase in the interface roughness scattering. Consequently, the electron velocity decreases as the polarization field increases.

4. Conclusions

In this paper, we have presented calculations of the electron drift velocity in $\text{Al}_x\text{Ga}_{1-x}\text{N}/\text{GaN}$ heterostructure devices in the presence of spontaneous and piezoelectric polarization fields. The calculations are made using

a comprehensive 2D ensemble Monte Carlo simulator. Fully numerical 2D scattering rates are incorporated into the Monte Carlo codes, based on the self-consistent solution of the Schroedinger and Poisson equations. We have introduced a new approach to finding the final state following a 2D polar optical scattering event. It is found that the usage of this new final state selection for polar optical phonon scattering is critical to properly determine the electron velocity in the 2D system. Based on this scheme our model provides an accurate description of electron velocity in HFET structures.

Acknowledgments

This work was supported in part by ONR through contract E21-K19, by NSF through grant ECS-9811366 and YAMACRAW.

References

- Brennan K.F. and Park D.H. 1989. *J. Appl. Phys.* 65: 1156.
- Farahmand M., Garetto C., Bellotti E., Brennan K.F., Goano M., Ghillino E., Ghione G., Albrecht J.D., and Ruden P.P. 2001. *IEEE Trans. Electron Dev.* 48: 535.
- Gaylord T.K. and Brennan K.F. 1989. *J. Appl. Phys.* 65: 814.
- Goano M., Bellotti E., Ghillino E., Ghione G., and Brennan K.F. 2000a. *J. Appl. Phys.* 88: 6467.
- Goano M., Bellotti E., Ghillino E., Ghione G., and Brennan K.F. 2000b. *J. Appl. Phys.* 88: 6476.
- Hsu L. and Walukiewicz W. 2001a. *J. Appl. Phys.* 89: 1783.
- Kawamura T. and Das Sarma S. 1992. *Phys. Rev. B* 45: 3612.
- Li T., Joshi R.P., and Fazi C. 2000. *J. Appl. Phys.* 88: 829.
- Park D.H. and Brennan K.F. 1989. *J. Appl. Phys.* 65: 1615.
- Ridley B.K. 1997. *Electrons and Phonons in Semiconductor Multi-layers*. Cambridge University Press, Cambridge.
- Tomizawa K. 1993. *Numerical Simulation of Submicron Semiconductor Devices*. Artech House, Boston.
- Yu T-H. and Brennan K.F. 2001b. *J. Appl. Phys.* 89: 3827.
- Yokoyama K. and Hess K. 1986. *Phys. Rev. B* 33: 5595.



Parallel Approaches for Particle-Based Simulation of Charge Transport in Semiconductors

M. SARANITI* AND J. TANG

Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL, USA

saraniti@iit.edu

S. GOODNICK AND S. WIGGER

Electrical Engineering Department, Arizona State University, Tempe, AZ, USA

Abstract. The aim of this contribution is to discuss possible algorithmic choices and hardware configurations for the implementation of efficient particle-based simulation programs. By using a population decomposition scheme, we modified the scalar version of the algorithm in order to improve the efficiency of our hybrid particle-based simulation engine. Using a Beowulf-class computer cluster, we measured parallel speed-up with different algorithmic configurations, and related it to the inter-process communication hardware.

Keywords: particle-based methods, parallel computing, charge transport simulation, Monte Carlo

1. Introduction

Since the early theoretical work on the Ensemble Monte Carlo (EMC) method applied to semiconductor simulation (Canali *et al.* 1975, Jacoboni and Reggiani 1983), and several subsequent reference books addressing both the physics and the numerical aspects of the EMC method (Hockney and Eastwood 1988, Jacoboni and Lugli 1989), the basic algorithmic approaches have been modified to exploit the continuous improvements of both hardware and software tools. In particular, the introduction of the full-band representation of the electronic structure (Fischetti and Laux 1988) and of the phonon dispersions, as well as the availability of fast Poisson solvers (Saraniti *et al.* 1996), extended the use of the particle-based code from a purely academic environment to the industrial laboratories. The extreme accuracy and numerical stability of particle-based simulation algorithms promoted the development of commercial versions of the research programs. However, the intrinsic complexity of the algorithms influences

the performance of the simulators, which require impractically long simulation times. Several algorithmic improvements have been steadily suggested within the last decade to optimize the use of the impressively improving computing hardware.

Algorithms have been first modified to take advantage of vector processing (Ravaioli 1991), and, more recently, the availability of large amounts of directly addressable random access memory (RAM) allowed the achievement of impressive speed-up by storing the complete transition table for all states in momentum space. This latter algorithmic development generated the “Cellular Monte Carlo”¹ (CMC) code, which is physically equivalent but up to 50 times faster than the traditional EMC (Saraniti and Goodnick 2000).

Furthermore, the partially local nature of some of the charge carrier interactions encouraged researchers in designing parallel variants of the basic algorithms. This development has been made possible by the availability of a relatively new class of parallel platforms: the computer cluster, based on standard networked workstations and on efficient and reliable inter-process communication software (Carns *et al.* 1999).

*To whom correspondence should be addressed.

In this paper, we will discuss the algorithmic approaches used to improve the efficiency of the particle-based algorithms running on such workstation clusters.

2. Hardware and Software Configuration

Any algorithmic configuration must be chosen by carefully considering the nature of the computing equipment used to run the parallel code. The computer used in this work is a Beowulf-class cluster of dual-processor nodes, each one equipped with 850 MHz processors and 2 Gbytes of RAM. The communication backbone is based on 100 Mbit Ethernet links, connected to a network switch. Clearly, the weak point of this hardware configuration is the communication channel, while the computing power of each individual processor is adequate. The choice of such a "slow" communication link was suggested by the availability of similar equipment in the academic world.

Inter-process communications is supplied by the Argonne National Laboratory implementation of the Message Passing Interface specifications (MPI) (Carns *et al.* 1999). Being based on the "message passing" paradigm, MPI seemed the most appropriate choice in terms of flexibility and reliability. A simple set of low-level binary communication functions was built on the top of MPI and configured, as usual, in an independent software layer. Besides the features present in the MPI implementation, no check on communication integrity has been implemented in the software layer. Since the processors in the cluster use the same internal number representation, no format conversion was necessary. Also, no real-time data compression techniques have been used in the MPI driver, assuming that the resulting compression ratio would be very low because of the type-less, binary nature of the transmitted data.

3. Algorithmic Structure

The typical algorithmic structure of a particle-based simulation program for device simulation is shown in Fig. 1. From an algorithmic viewpoint, the sequential nature of the scheme is evident (Hockney and Eastwood 1988, Jacoboni and Lugli 1989). In fact, the requirement of self-consistence between carrier dynamics and electric field implies (1) the need of a synchronous carrier ensemble (Fischetti and Laux 1988), and (2) some sort of efficient approach to the solution of Poisson's equation. This work is mostly concerned about the first requirement, while the problem of how to distribute

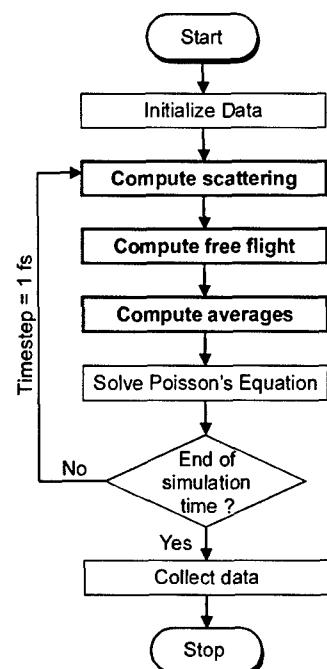


Figure 1. Flowchart of a typical particle-based simulation algorithm.

the computational load due to the Poisson solver will be only briefly discussed.

The approach chosen in this research is to find a decomposition of the carrier population that allows for satisfactory speed-up, while keeping the components of the system synchronous. In particular, information about the total charge distribution has been made available to an individual process at the end of any iteration, ready for the Poisson solver. This strategy only requires the modification of the few algorithmic modules that are used to update the dynamics of the carriers during the simulation, and are depicted in bold in the flowchart of Fig. 1.

The approach chosen in order to share the computational load is the so-called population decomposition method, and it is performed by splitting the carrier population and by assigning a portion of it to each concurrent process. Keeping in mind the need of a synchronous ensemble, we tested the performance of the two different algorithmic configurations.

The first algorithm, shown in Fig. 2, has been designed to minimize the inter-process communication flow. The basic idea is that, while the free-flight algorithm changes the status of all carriers at any iteration, only a small portion of them are subject to scattering. Basing on this observation, the algorithm in

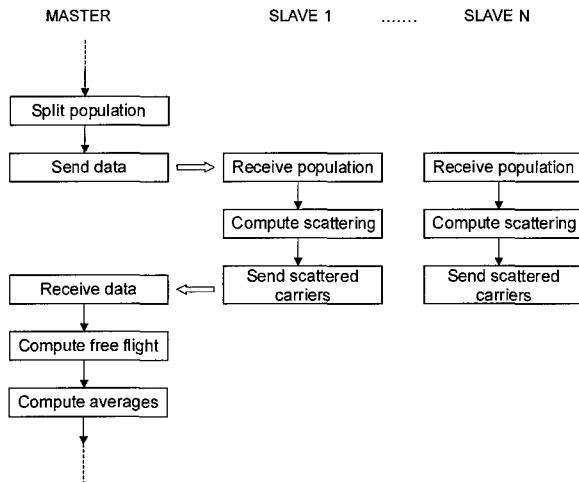


Figure 2. Flowchart of the parallel procedure with scattering sharing.

Fig. 2 shares only the scattering computation between processes, while devoting one processor (the so called “master” processor) to data collection and to the free-flight calculations. While relatively unbalanced, this algorithm is based on the fact that the larger computational burden of an EMC code is due to the scattering computation (Saraniti and Goodnick 2000). Unfortunately, the very short communication time typical of this algorithm does not generate the expected speed-up. Simulation results are shown in Fig. 3, where saturation of the speed-up is noticeable, and the overall performance improvement is definitely unsatisfactory. Even if it minimizes communication, the algorithm in Fig. 2 generates a load unbalance responsible for the performance degradation. Improved parallel speed-up can be achieved by including the free-flight computation in the shared part of the algorithm, as shown in Fig. 4.

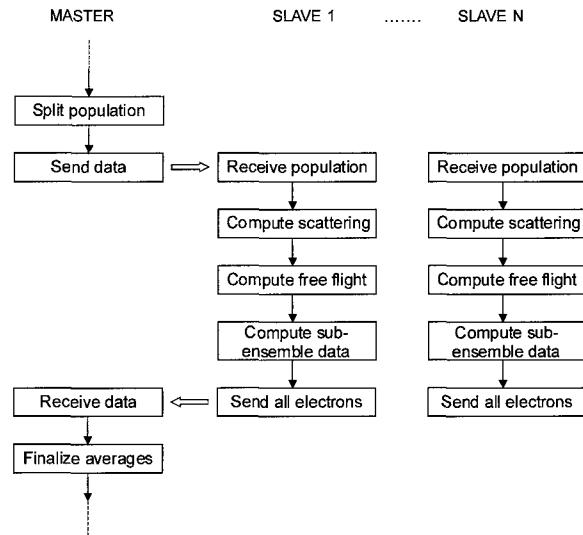


Figure 4. Flowchart of the parallel procedure with scattering and free-flight sharing.

In this case, both the free-flight calculations and some of the averaging are tasks performed by the “slave” processes, while the “master” process takes care of collecting data and distributing the carrier population. This choice leaves the master node mostly inactive during the simulation, leaving CPU resources free for the inclusion of a Poisson solver. In fact, preliminary estimates show that, for a typical situation, the time to solve Poisson’s equation in two dimensions is comparable to the time required by the slave “processes” to process scattering and free-flight. The parallel performance of the algorithm of Fig. 4 is depicted in Fig. 5.

The speed-up for seven slave processes is still sub-linear, but a remarkable improvement has been achieved with the new algorithm. Also, simulation

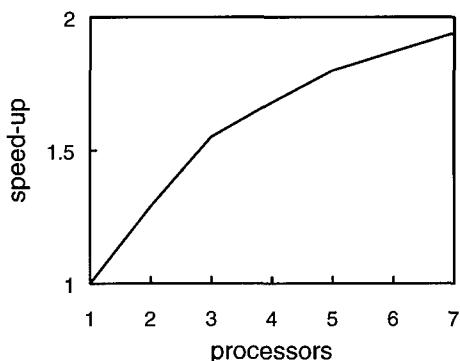


Figure 3. Parallel performance of the algorithm in Fig. 2.

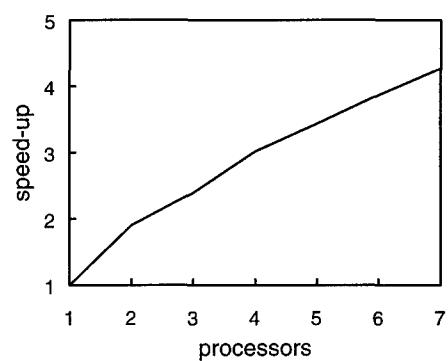


Figure 5. Parallel performance of the algorithm in Fig. 4.

data clearly shows that the computational load between slave processes is well balanced, and that the performance can be significantly improved only by minimizing the communication time. This can be obtained by using a faster local area network. Commercially available LAN's can achieve inter-process transfer rates up to 20 times faster than the Ethernet link used in the shown simulations.

4. Result Validation

Results obtained with the parallel procedures described above were validated with experimental results, and with scalar runs. In particular, both steady-state and transient response of electrons in Gallium Arsenide (GaAs) were simulated, and the main quantities compared with experiment. The speed-up evaluations shown in Figs. 3 and 5 were obtained with an EMC simulation of an ensemble of 5000 electrons in GaAs with an external field of 5×10^6 V/m.

5. Conclusions

We have performed evaluations of parallel performance of particle-based algorithms on a computer cluster using the message-passing paradigm for inter-process communication. Both load-balancing and communication time play a crucial role in the parallel performance of the code. Simulation results clearly show the possibility of achieving a more than satisfactory speed-up, especially if high performance communication hardware is used.

Acknowledgments

This work has been partially supported by the National Science Foundation Grant ECS-9976484, and by the

ERIF fund of the Illinois Institute of Technology. The authors would like to acknowledge the contribution of Julien Branlard in the simulation and editing work.

Note

1. The approach was initially called "cellular automaton", and evolved in an algorithmic structure that shares few, if any, of the basic aspects of the homonym method used in computational hydrodynamics. For this reason the "Cellular Monte Carlo" denomination was recently introduced.

References

- Canali C., Jacoboni C., Nava F., Ottaviani G., and Alberigi Quaranta A. 1975. Electron drift velocity in silicon. *Physical Review B* 12(4): 2265.
- Carns P.H., Ligon W.B., McMillan S.P., and Ross R.B. 1999. An evaluation of message passing implementations on Beowulf workstations. In: *Proceedings of Aerospace Conference*, Vol. 5, p. 41.
- Fischetti M.V. and Laux S. 1988. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects. *Physical Review B* 38(14): 9721.
- Hockney R.W. and Eastwood J.W. 1988. *Computer Simulation using Particles*. Adam Hilger, Bristol.
- Jacoboni C. and Lugli P. 1989. *The Monte Carlo Method for Semiconductor Device Equations*. Springer-Verlag, Wien.
- Jacoboni C. and Reggiani L. 1983. The Monte Carlo method for solution of charge transport in semiconductors with applications to covalent materials. *Reviews of Modern Physics* 55(3): 645.
- Ravaioli U. 1991. Vectorization of Monte Carlo algorithms for semiconductor simulation. In: Hess K. (Ed.), *Monte Carlo Device Simulation: Full Band and Beyond*. Kluvier, Dordrecht, p. 267.
- Saraniti M. and Goodnick S.M. 2000. Hybrid full-band cellular automaton/Monte Carlo approach for fast simulation of charge transport in semiconductors. *IEEE transactions on Electron Devices* 47(10): 1909.
- Saraniti M., Rein A., Zandler G., Vogl P., and Lugli P. 1996. An efficient multigrid Poisson solver for device simulations. *IEEE Transaction on Computer-Aided design of Circuits and Systems* 15(2): 141.



Full-Band Monte Carlo Simulation of Two-Dimensional Electron Gas in SOI MOSFETs

H. TAKEDA,* N. MORI AND C. HAMAGUCHI

Department of Electronic Engineering, Osaka University, 2-1 Yamada-oka, Suita City, Osaka 565-0871, Japan

takeda@ele.eng.osaka-u.ac.jp

Abstract. A full-band Monte Carlo simulation of two-dimensional electron gas is performed to study effects of the non-parabolicity of the energy band structure on the phonon-limited electron mobility in SOI MOSFETs with a thin Si-layer.

Keywords: SOI MOSFET, pseudo-potential method, energy band structure, Monte Carlo simulation

1. Introduction

The full-band Monte Carlo technique is a powerful tool for simulating carrier motion in bulk materials. It has been reported that a full-band Monte Carlo model for bulk silicon gives a good quantitative agreement with the temperature, electric field, and crystal direction dependences of experimental electron and hole drift-velocities (Fisher and Hofmann 2000). In metal-oxide-semiconductor (MOS) structures, however, the electron motion in the direction perpendicular to the surface is quantized by the strong electric field applied across the interface, and this two-dimensional quantization can strongly affect the carrier transport properties. It has been shown that, for example, silicon-on-insulator metal-oxide-semiconductor field effect transistors (SOI MOSFETs) with thinner Si-layer than the inversion layer of the bulk MOSFETs can provide higher electron mobility than the bulk Si-MOSFETs because of the significant modulation of the subband structure due to the quantization effects (Takagi, Koga and Toriumi 1998). It is, therefore, desirable to develop a Monte Carlo code capable of simulating the transport properties of two-dimensional electron gas (2DEG) with full-band modeling. In the present study, we have performed a full-band Monte Carlo simulation of 2DEG and studied effects of the non-parabolicity of

the energy band structure on the transport properties in SOI MOSFETs with a thin Si-layer.

2. Self-Consistent Calculation

The two-dimensional electronic states in SOI MOSFETs were calculated by a self-consistent method using an empirical pseudo-potential approximation (Takeda, Mori and Hamaguchi 2001). For the pseudo-potential $V_P(\mathbf{R})$ of bulk silicon, the Schrödinger equation for SOI MOSFETs with a confining potential $U(z)$ along the z -direction, the direction perpendicular to the surface, may be written as

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V_P(\mathbf{R}) + U(z) \right\} \Psi(\mathbf{R}) = E \Psi(\mathbf{R}). \quad (1)$$

By expanding the wave function $\Psi(\mathbf{R})$ with the bulk wave functions of $\psi_{\mathbf{K}}^n(\mathbf{R}) = e^{i\mathbf{K}\cdot\mathbf{R}} u_{\mathbf{K}}^n(\mathbf{R}) = e^{i\mathbf{K}\cdot\mathbf{R}} \sum_{\mathbf{G}} f_{\mathbf{K}}^n(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{R}}$, we have

$$\begin{aligned} & \sum_n \sum_{\mathbf{G}'} \sum_{k_z, g_z} c_{\mathbf{k}', k_z}^n f_{\mathbf{K}'}^{n'*}(\mathbf{G}') \\ & \times f_{\mathbf{k}', k_z}^n(\mathbf{g}', g_z) U_F(k_z - k_z' + g_z - g_z') \\ & = \sum_n \sum_{\mathbf{G}'} [E - \mathcal{E}^n(\mathbf{K}')] c_{\mathbf{k}', k_z}^n f_{\mathbf{K}'}^{n'*}(\mathbf{G}') f_{\mathbf{K}}^n(\mathbf{G}'), \quad (2) \end{aligned}$$

where $c_{\mathbf{k}', k_z}^n$ are the expansion coefficients ($\Psi(\mathbf{R}) = \sum_{\mathbf{K}, n} c_{\mathbf{K}}^n \psi_{\mathbf{K}}^n(\mathbf{R})$), $\mathcal{E}^n(\mathbf{K})$ the energy band of bulk

*To whom correspondence should be addressed.

silicon, n the band index, \mathbf{K} the three-dimensional wave vector, and

$$U_F(q_z) = \frac{1}{L} \int_0^L U(z) e^{iq_z z} dz \quad (3)$$

with L being the thickness of the system. In the following calculation, we assume that $U(z)$ is slowly varying and $U_F(k_z - k'_z + g_z - g'_z)|_{g_z \neq g'_z} = 0$.

To obtain the confining potential $U(z)$, we solved the Poisson equation within the Hartree approximation, which requires electron charge distribution along the z direction, $\rho_e(z)$. We evaluated $\rho_e(z)$ using the Fermi-Dirac distribution function and state densities calculated with discretizing the (k_x, k_y) -space into rectangle meshes (Takeda, Mori and Hamaguchi 2001).

For studying effects of the non-parabolicity, we also performed self-consistent calculation using the effective mass approximation with a simple parabolic energy band.

3. Scattering Rates

For a Monte Carlo simulation, we calculated the electron scattering rates. In the present study, we include intra-valley scatterings via acoustic phonons and inter-valley scatterings via acoustic or optical phonons with arbitrarily defining the boundary of the valleys (Fischetti and Laux 1988). The scattering rates are evaluated at each mesh point in the discretized (k_x, k_y) -space, which is used for evaluating the state densities in the self-consistent calculation.

The intra-valley scattering rate via acoustic phonons from $\mathbf{k} (= (k_x, k_y))$ in subband μ is given by

$$\frac{1}{\tau_{\text{intra}}^\mu(\mathbf{k})} = \frac{k_B T}{\hbar \rho v_s^2} \sum_v \sum_{k'_x, k'_y} G_{k'_x, k'_y}^v(E^\mu(\mathbf{k})) \times \int \Delta^2(\mathbf{k}' - \mathbf{k}, q_z) |\mathcal{F}_{\mu, v}(q_z)|^2 dq_z, \quad (4)$$

where q_z is the z -component of the three-dimensional phonon wave vector \mathbf{Q} , ρ the density of silicon, v_s the sound velocity, $E^\mu(\mathbf{k})$ the energy eigenvalue of Eq. (2) and $G_{k'_x, k'_y}^\mu(E^\mu(\mathbf{k}))$ the state densities. In Eq. (4), both the absorption and the emission processes are included within the equipartition approximation. We take into account longitudinal acoustic (LA) and transverse acoustic (TA) phonons. The anisotropic deformation potential $\Delta(\mathbf{k})$ for each phonon mode is given by the

following equations

$$\Delta_{\text{LA}}(\theta_Q) = \Xi_d + \Xi_u \cos^2(\theta_Q), \quad (5)$$

$$\Delta_{\text{TA}}(\theta_Q) = \Xi_u \cos^2(\theta_Q) \sin(\theta_Q), \quad (6)$$

where $\Xi_u = 9.0 \text{ eV}$, $\Xi_d = -11.7 \text{ eV}$ and θ_Q is the angle between phonon wave vector \mathbf{Q} and the longitudinal axis of each valley (Fischetti and Laux 1993). The form factor $\mathcal{F}_{\mu, v}(q_z)$ is defined by the following equation:

$$\mathcal{F}_{\mu, v}(q_z) = \int \Psi_v^*(z) e^{iq_z z} \Psi_\mu(z) dz. \quad (7)$$

The inter-valley phonon scattering rate is given by

$$\begin{aligned} \frac{1}{\tau_{\text{inter}}^\mu(\mathbf{k})} &= \frac{\Delta^2}{2\rho\omega_0} \sum_v \int dq_z |\mathcal{F}_{\mu, v}(q_z)|^2 \\ &\times \sum_{k'_x, k'_y} [N_0 G_{k'_x, k'_y}^v(E^\mu(\mathbf{k}) + \hbar\omega_0) \\ &+ (N_0 + 1) G_{k'_x, k'_y}^v(E^\mu(\mathbf{k}) - \hbar\omega_0)], \end{aligned} \quad (8)$$

with $N_0 = (e^{\hbar\omega_0/k_B T} - 1)^{-1}$ being the phonon population number. We include g -processes via TA, LA, and longitudinal optical (LO) phonons, and f -processes via TA, LA, and transverse optical (TO) phonons. The values of the deformation potential Δ and the phonon energy $\hbar\omega_0$ of each phonon mode are listed in Table 1.

Figure 1 shows the total scattering rates of the lowest subband in the 2-fold valleys as a function of the electron energy in a SOI MOSFET with the Si-layer thickness of $L_{\text{Si}} = 5.43 \text{ nm}$ at $T = 300 \text{ K}$ for the voltage difference between the Si-layer of $\Delta V_{\text{Si}} = 0.1 \text{ V}$. The closed circles represent the total scattering rate calculated with the pseudo-potential model, and the solid

Table 1. Parameters used in the present calculation (Fischetti and Laux 1993).

	Deformation potential (10^{10} eV/m)	Phonon energy (meV)
TA-phonons (g -process)	0.5	12
LA-phonons (g -process)	0.8	18.5
LO-phonons (g -process)	11.0	61.2
TA-phonons (f -process)	0.3	19.0
LA-phonons (f -process)	2.0	47.4
TO-phonons (f -process)	2.0	59.0

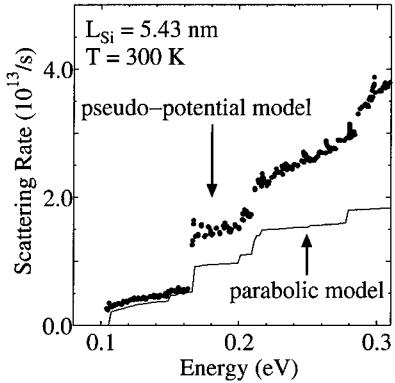


Figure 1. Total scattering rates of the lowest subband in the 2-fold valleys calculated with the pseudo-potential (closed circles) and the parabolic (solid lines) models.

line is that for the parabolic model. For higher electron energy than ≈ 0.15 eV, we see that the difference between the two models gradually increases as the electron energy increases.

4. Monte Carlo Simulation

Using the scattering rates calculated in the previous section, we performed a single-electron Monte Carlo simulation. Figure 2 shows L_{Si} -dependence of the low-field mobilities of SOI MOSFETs with the electron sheet density of $N_s = 4.0 \times 10^{12} \text{ cm}^{-2}$ at $T = 300$ K for the voltage difference between the Si-layer of $\Delta V_{\text{Si}} = (L_{\text{Si}}/5.43 \text{ nm}) \times 0.1 \text{ V}$ (i.e. $F_{\perp}^{\text{ave}} = 0.18 \text{ MV/cm}$). The parallel electric field is applied along the x -direction ($\mathbf{F}_{\parallel} = (F_x, 0)$) with $F_x \approx 1 \text{ kV/cm}$. L_{Si} -dependence of

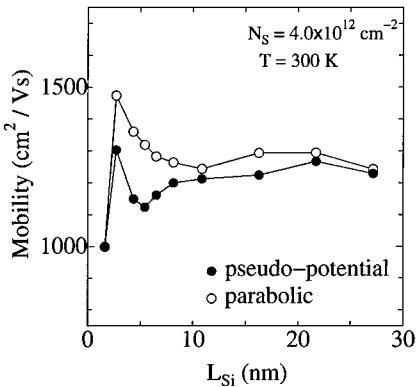


Figure 2. Calculated low-field phonon-limited mobility as a function of the Si-layer thickness. The results of the pseudo-potential model are plotted as the closed circles and those of the parabolic model are plotted as the open circles.

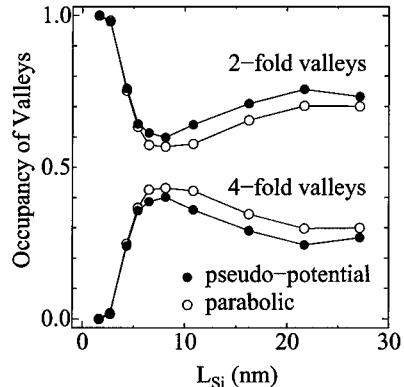


Figure 3. The occupancy of the 2-fold and the 4-fold valleys as a function of the Si-layer thickness. The closed and open circles represent the results of the pseudo-potential and the parabolic models, respectively.

occupancy of the 2-fold and the 4-fold valleys is also plotted in Fig. 3.

For $L_{\text{Si}} \gtrsim 10$ nm, the occupancy of the 2-fold valleys of the pseudo-potential model is larger than that of the parabolic model because of the larger state densities of the pseudo-potential model, which is originated in the non-parabolicity of the conduction band structure. Although the scattering rates are higher in the pseudo-potential model (see Fig. 1), the low-field mobilities do not differ so much between the two models for $L_{\text{Si}} \gtrsim 10$ nm. This is because the larger occupancy of the 2-fold valleys, where the electron mobility is higher than that in the 4-fold valleys, partially compensates the higher scattering rate in the pseudo-potential model.

For $L_{\text{Si}} \lesssim 10$ nm, the subband spacing increases as L_{Si} decreases both in the two models, resulting in an increase in the occupancy of the 2-fold valleys (see Fig. 3). The difference of the occupancy between the two models then practically vanishes for $L_{\text{Si}} < 5$ nm. This leads to a large difference in the low-field mobility between the two models for $L_{\text{Si}} \lesssim 5$ nm as shown in Fig. 2. In the limit of the vanishing L_{Si} , the scattering rates diverge and the difference in the electron mobility disappears as $L_{\text{Si}} \rightarrow 0$.

5. Summary

We solved the Schrödinger and the Poisson equations self-consistently within the empirical pseudo-potential and the Hartree approximations to obtain the two-dimensional electronic states in SOI MOSFETs.

We then carried out the Monte Carlo simulation using the scattering rates evaluated with the electronic states obtained by the self-consistent calculation. We find that the non-parabolicity of the conduction band structure significantly influences the phonon-limited electron mobility for $L_{\text{Si}} \approx 5 \text{ nm}$.

References

- Fisher B. and Hofmann K.R. 2000. *Appl. Phys. Lett.* 76: 583.
Fischetti M.V. and Laux S.E. 1988. *Phys. Rev. B* 38: 9721.
Fischetti M.V. and Laux S.E. 1993. *Phys. Rev. B* 48: 2244.
Takagi S., Koga J., and Toriumi A. 1998. *Jpn. J. Appl. Phys.* 37: 1289.
Takeda H., Mori N., and Hamaguchi C. 2002. *Physica B* 314: 377.



Band-to-Band Tunneling by Monte Carlo Simulation for Prediction of MOSFET Gate-Induced Drain Leakage Current

EDWIN C. KAN, VENKAT NARAYANAN AND GEN PEI

School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA

Abstract. Gate induced drain leakage (GIDL) current caused by band-to-band tunneling is studied by Monte Carlo simulation with ballistic least-action trajectory integration. Together with weak inversion and early subthreshold simulation by drift-diffusion formalism, the entire range of the OFF-state drain current can be predicted for technology evaluation. The methodology is demonstrated by a case study for source/drain asymmetry super-halo design.

Keywords: MOSFET, leakage current, band-to-band tunneling, Monte Carlo simulation, source-drain asymmetry

1. Introduction

Aggressive scaling of CMOS technology below $0.1 \mu\text{m}$ gate pitch has stringent requirements on standby power consumption and parametric yield. MOSFET OFF-state drain current, mainly affected by the subthreshold slope and the drain junction tunneling leakage (by doping or by gate-induced accumulation), is a key device parameter in early technology evaluation. Due to thin-oxide direct tunneling leakage and gate capacitive loading, oxide scaling cannot be as aggressive below $0.1 \mu\text{m}$ technology (Hu 1996). To control the short-channel effects (SCE) for higher parametric yield, substrate solutions such as the super-halo doping and SOI structures become strong candidates in device design (Taur and Nowak 1997, Wann *et al.* 1996, Fossum, Kim and Chong 1999). Accurate prediction of the OFF-state drain current for these structures requires detailed *multi-dimensional distributive* treatment of potential profiles and tunneling probabilities, where simplified analytical formations (Fossum, Kim and Chong 1999, Jomaah, Ghibaudo and Balestra 1996) are not sufficiently predictive. We propose a hybrid simulation method that can take the distributive effects rigorously for leakage prediction. The part of gate bias sweep containing flat band and weak inversion is treated by the drift-diffusion formalism, since neither quantum (no channel confinement potential) nor nonequilibrium (low drain current) effect is critical.

The part of gate sweep driving the channel to deep accumulation where we anticipate a significant band-to-band tunneling is modeled by Monte Carlo simulation (Laux and Fischetti 2000), where the detailed momentum information is available for “steepest-decent” ballistic trajectory path integration (Tanaka, Sasaki and Yamamoto 1994). As an illustrative example, the proposed method is applied to study the leakage currents in source/drain asymmetric super-halo structures (Kan, Jeong and Wong 2001), where the effect of band-to-band tunneling is essential to design the super-halo doping profile.

2. Indirect Band-to-Band Tunneling Model

The model used for the Monte Carlo simulation (Laux and Fischetti 2000) is based on the indirect band-to-band tunneling current density expressions derived by Tanaka (1994), with improvement to include non-constant electric fields and least-action trajectory search (Fischetti 2000). The approach in deriving these expressions is to start from the Wannier equation in 1-D (1).

$$[E_n(\hat{k}) + U(x)]\psi_n(r, t) + \sum_{n \neq n'} W_{nn'}\psi_{n'}(r, t) - i\hbar \frac{\partial \psi_n(r, t)}{\partial t} = 0 \quad (1)$$

where $U(x)$ is the potential energy and $\psi_n(r, t)$ is the electron wave function in the n band. This is just a re-statement of the time-dependent Schrödinger equation with the coupling terms $W_{nn'}$ between various bands n and n' explicitly included. The operator form of the wave vector k is used in the energy term $E_n(k)$ and determines the order of the differential equation.

This equation is first solved in real-space in the absence of the driving force for tunneling (the non-diagonal interband coupling term) for a two-band model and the wave functions for the conduction and valence bands are obtained. For indirect bandgap materials like silicon, the coupling term in (1) is due to the phonons since a large crystal momentum change is required for the tunneling transition. Time dependent perturbation theory in the form of the Fermi-Golden rule is used to calculate the transition probability between the states. This 1-D model has been implemented in a trajectory based quasi 2-D model in Laux and Fischetti (2000), through a search for local electric fields large enough to cause a generation event (Fischetti 2000). The generation rates are separately calculated for the conduction band valleys and the valence bands to generate the carriers at the correct regions of the k -space. Notice that in a partial-differential-equation (PDE) based device simulator, the band-to-band tunneling can only be estimated by perturbed equilibrium distribution, which may introduce significant errors for indirect tunneling in the highly nonequilibrium situation, such as the drain depletion region in consideration of GIDL in the next section.

3. Case Study on Source/Drain Asymmetrical Super-Halo Design

One of the possible ways to create sub-50 nm bulk CMOS devices is through the use of super-halo doping (Taur and Nowak 1997). Since the channel doping mainly originates from tilted implantation through self-aligned source/drain extension, an additional mask can create selective source or drain doping. This S/D asymmetry has been proposed to be used to improve circuit performance (Kan, Leong and Wong 2001, Buti *et al.* 1991, Kumagai *et al.* 1994, Hikori, Odanaka and Hori 1995, Ohzone *et al.* 1997) based reduced parasitics or difference in threshold voltage in forward (V_{thF}) and backward (V_{thR}) situations, which is created by the different DIBL (drain induced barrier lowering) effects on source/drain reversal. DIBL is more effective when the source end potential has a smaller curva-

ture (net charge density or looser potential boundary conditions). For a subthreshold slope of 80–100 mV, the current ratio between reasonable $|V_{thF} - V_{thR}|$ can be as large as 500–5000 times (depending on the noise margin given for above-threshold). However, the heavy super-halo implant can enlarge band-to-band tunneling, which is especially important in the GIDL case. Doping implantation needs to be optimized to *adjust the channel implant profiles* to minimize the band-to-band tunneling while maintaining tight control on SCE.

The device geometry used here is similar to Yang *et al.* (1999) with super-halo doping design follows the proposal in Taur and Nowak (1997). The device geometry and net doping profiles are shown in Fig. 1. The Poisson equation dictates that net charge is the potential curvature. At subthreshold V_{GS} and high saturation V_{DS} , clearly no channel carriers are significant and the only net charge is from the doping, i.e., only high doping regions can possibly have high potential curvature. For the source-side super halo, the source barrier can be maintained from this curvature, while the drain-side super-halo has a much lower source barrier with a built-in field toward the drain. This is the main physical reason for the difference between V_{thF} and V_{thR} . Drift-diffusion simulation is used to demonstrate the super-halo doping effect on the threshold voltage and SCE behavior in Fig. 2. Four structures are compared, super-halo at both source and drain sides, at only source or drain side, and no super-halo where punchthrough can be observed.

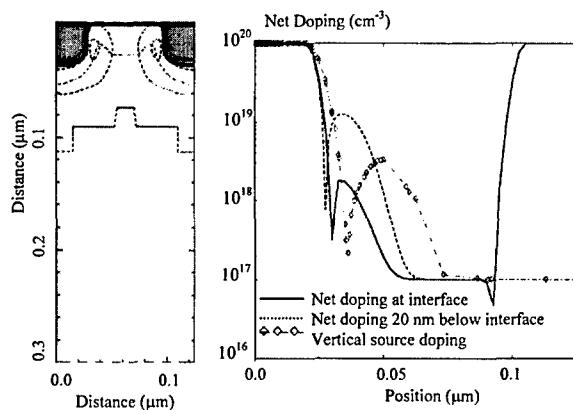


Figure 1. 2D Net doping profile in the super-halo MOSFET is shown at left. 1D cross section for the net doping profiles at right. Solid line: cut line at the Si/SiO₂ interface; dash line: parallel profile at 20 nm below the Si/SiO₂ interface; diamonds: vertical profile in the source region.

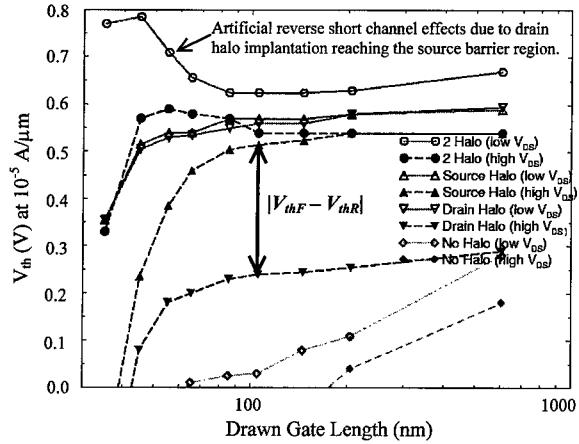


Figure 2. Short channel effects (SCE) in the super-halo bulk CMOS devices. Calculations are based on 2D drift-diffusion simulation. One-sided super-halo design can achieve acceptable SCE behavior below 70 nm drawn gate length.

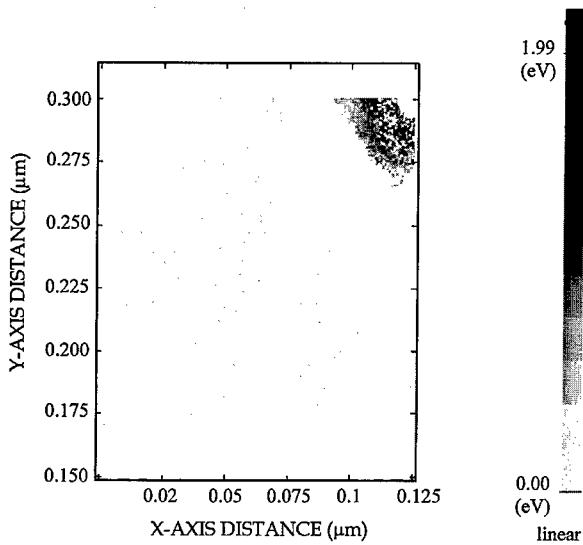


Figure 3. Monte Carlo simulation of electrons under above-threshold gate bias and high drain bias in drain-halo configuration. The bias condition is posed such that the drain halo region is strong enough to cause potential isolation in the substrate, but the drain-induced-barrier lowering is significant at the interface.

Figure 3 shows the electrons and their energy in the Monte Carlo simulation, while Fig. 4 is the generation rate estimated from the band-to-band tunneling calculated in the Monte Carlo simulation. Figure 5 demonstrates determination of the achievable lowest leakage current using the hybrid method. When the gate is swept to deep subthreshold region, the tunneling leakage cur-

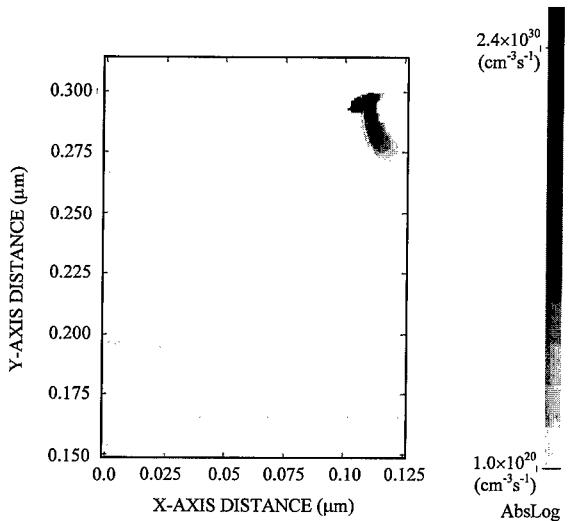


Figure 4. Generation rate from band-to-band tunneling is obtained from Monte Carlo simulation corresponding to the same device in Fig. 3. V_{GS} is in deep subthreshold and V_{DS} is high. The generation rate is highest in the drain junction close to the interface, but is nonuniform.

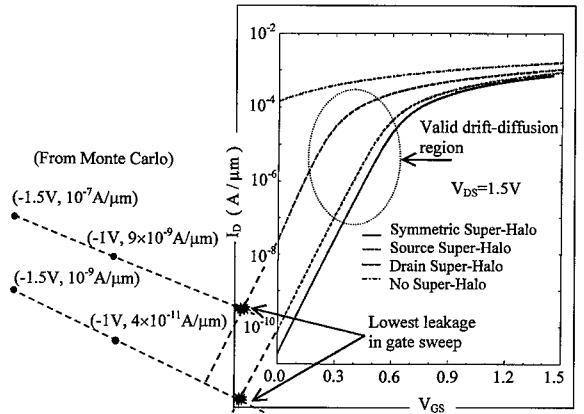


Figure 5. Extrapolation for achievable lowest leakage from combined drift-diffusion and Monte Carlo analyses. Drain super halo design will not only have stronger DIBL, but also larger band-to-band tunneling due to the potential distribution.

rent is large enough that the statistical fluctuation in Monte Carlo is tolerable in the log-scale plot. Notice that the actual experimental measurement may not be able to use this strategy to delineate the tunneling leakage easily due to the limited oxide breakdown field by V_{GD} . Drift-diffusion formalism is also inaccurate due to the lack of k -space information and effective cross section estimation. These points are used to

extrapolate on the drift-diffusion subthreshold (remember that DD is accurate in the shallow subthreshold region) to obtain the theoretically lowest I_{OFF} of the MOSFET under study. The drain-side halo has higher band-to-band tunneling as expected, but the lowest I_{OFF} is still in the acceptable range $4 \times 10^{-10} \text{ A}/\mu\text{m}$ for the doping design in hand. The proposed hybrid method is shown to be effective for technology evaluation on substrate solutions in the early phase of process development.

Acknowledgments

This project originated from IBM summer faculty partner program in 2000 for E. C. K., and was later developed under the support of SRC 848.001. Discussion and support from S. Laux, M. Fischetti, M. Ieong, H.-S.P. Wong and P. Oldiges (also the SRC Liason) of IBM are appreciated. Special thanks to M. Fischetti for his generous provision of unpublished band-to-band tunneling models in DAMOCLES 3.0.

References

- Buti T.N., Ogura S., Rovedo N., and Tobimatsu K. 1991. A new asymmetrical halo source GOLD drain (HS-GOLD) deep sub-half-micrometer n-MOSFET design for reliability and performance. *IEEE Trans. Electron Devices* 38(8): 1757–1764.
- Fischetti M.V. 2000. Private communication.
- Fossum J.G., Kim K., and Chong Y. 1999. Extremely scaled double-gate CMOS performance projections, including GIDL-controlled off-state current. *IEEE Trans. Elec. Dev.* 46: 2195–2200.
- Hikori A., Odanaka S., and Hori A. 1995. A high-performance $0.1 \mu\text{m}$ MOSFET with asymmetric channel profile. *IEDM Tech. Dig.*, pp. 439–442.
- Hu C. 1996. Gate oxide scaling limits and projection. *IEDM Tech. Dig.*, pp. 319–322.
- Jomaah J., Ghibaudo G., and Balestra F. 1996. Band-to-band tunneling model of gate induced drain leakage current in submicron MOS transistors. *Electronics Letters* 32(8): 767–769.
- Kan E.C., Ieong M., and Wong P. 2001. Use of source/drain asymmetry MOSFET devices in dynamic and analog circuits. IBM Patent, Docket No. FIS920000389.
- Kumagai K., Kurosawa S., Iwaki H., Hamatake N., Yoshino A., Okumura K., Ohuchi K., Nakajima K., Asahina A., and Yamazaki Y. 1994. A mixed asymmetric/symmetric (MASS) MOSFET cell for ASICs. In: Proc. IEEE Intl. ASIC Conf. and Exhibit.
- Laux S.E. and Fischetti M.V. 2000. DAMOCLES Version 3.0. IBM Thomas Watson Research Lab.
- Ohzone T., Miyakawa T., Matsuda T., Yabu T., and Odanaka S. 1997. Performance evaluation of CMOS ring-oscillators with source/drain regions fabricated by asymmetric/symmetric ion implantation. In: IEEE Intl. Conf. Microelectronics Test Structures.
- Tanaka S. 1994. A unified theory of direct and indirect interband tunneling under a nonuniform electric field. *Solid State Electronics* 37(8): 1543–1552.
- Tanaka T., Sasaki M., and Yamamoto K. 1994. Field-theoretical description of quantum fluctuations in the multi-dimensional tunneling approach. *Physical Review D* 49: 1039–1046.
- Taur Y. and Nowak E.J. 1997. CMOS devices below $0.1 \mu\text{m}$: how high will performance go? *IEDM Tech. Dig.*, pp. 215–218.
- Wann C.H., Tu R., Yu B., Hu C., Noda K., Tanaka T., and Yoshida M. 1996. A comparative study of advanced MOSFET structures. In: Proc. Symp. VLSI Technology, pp. 32–33.
- Yang I.Y., Chen K., Smeys P., Sleight J., Lin L., Ieong M., Nowak E., Fung S., Maciejewski E., Varekamp P., Chu W., Park H., Agnello P., Crowder S., Assaderaghi F., and Su L. 1999. Sub-60 nm physical gate length SOI CMOS. *IEDM Tech. Dig.*



A Computational Technique for Electron Energy States Calculation in Nano-Scopic Three-Dimensional InAs/GaAs Semiconductor Quantum Rings Simulation

YIMING LI^{*,†}

National Nano Device Laboratories, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan; National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan

yml@cc.nctu.edu.tw

O. VOSKOBOYNIKOV

National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan; Kiev Taras Shevchenko University, 01033, Kiev, Ukraine

C.P. LEE

National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan

S.M. SZE

National Nano Device Laboratories, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan; National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan

Abstract. We study theoretically the electron energy states for three-dimensional (3D) nano-scopic semiconductor quantum rings. In this study, the model formulation includes: (i) the effective one-band Hamiltonian approximation, (ii) the position and energy dependent quasi-particle effective mass approximation, (iii) the finite hard wall confinement potential, and (iv) the Ben Daniel-Duke boundary conditions. To calculate the energy levels, the 3D model is solved by nonlinear iterative algorithm to obtain self-consistent solutions. The model and solution method provide a novel way to calculate the energy levels of nano-scopic semiconductor quantum ring and are useful to clarify the principal dependencies of quantum ring energy states on material band parameter, ring size and shape. We find the energy levels strongly depend on the radial cross section shapes of quantum rings. The dependence of energy states on shapes of 3D quantum ring reveals a significant difference from results derived on basis of 2D approaches.

Keywords: nano-scopic, semiconductor quantum rings, InAs/GaAs, energy states, computer simulation

1. Introduction

Recent progresses in the fabrication of semiconductor nanostructures make it possible to fabricate nano-scopic quantum rings with various geometries

(Bimberg *et al.* 2000, Emperador *et al.* 2000, Li *et al.* 2001a, b, Bastard 1988, Bruno-Alfonso and Latgé 2000, Lorke *et al.* 2000, Tsai *et al.* 1998). Although micro-scopic and meso-scopic metallic semiconductor quantum rings have been of a considerable attention in recent years, the development in fabrication of semiconductor nano-scopic rings significantly bridges the gap between quantum dots and meso-scopic quantum ring structures. Most theoretical quantum ring models

*To whom correspondence should be addressed.

†Present address: Microelectronics and Information Research Center, National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu 300, Taiwan.

assume only electrons moving in a 2D plane confined by a parabolic potential (Emperador *et al.* 2000, Li *et al.* 2001a).

These models, however, do not consider some important phenomena, such as (i) effect of the inner or outer radius of the ring, (ii) the finite hard wall confinement potential, and (iii) effect of non-parabolic band approximation for electron effective mass. Therefore, for a more comprehensive study a complete theoretical model for electron energy states in realistic 3D nanoscopic semiconductor quantum rings simulation should be taken into consideration. In addition, in this case the nonlinear eigenvalue problem cannot be solved exactly and a numerical simulation technique is required.

In this paper, the model is formulated and solved numerically based on the effective three-dimensional one band Hamiltonian, the energy (non-parabolic), the position and energy dependent quasi-particle effective mass approximation (Li *et al.* 2001b), and the Ben Daniel-Duke boundary conditions. We developed a nonlinear iterative method to solve the energy dependent Schrödinger equation and obtain the self-consistent results. A shifted and balanced QR algorithm as well as inverse iteration method is applied to compute the electron energy states and the corresponding wave functions. With the developed quantum ring simulator, a realistic (ring with ellipsoidal shape cross section) 3D model for InAs/GaAs quantum rings with the finite hard wall confinement potential simulation is simulated successfully. Quantum rings with rectangular and ellipsoid shape cross sections (see Fig. 1) are simulated and compared to show the significant variation in energy levels. Section 2 presents the 3D quantum ring model and computation algorithms. Section 3 demonstrates and discusses the simulation results. Section 4 draws the conclusions.

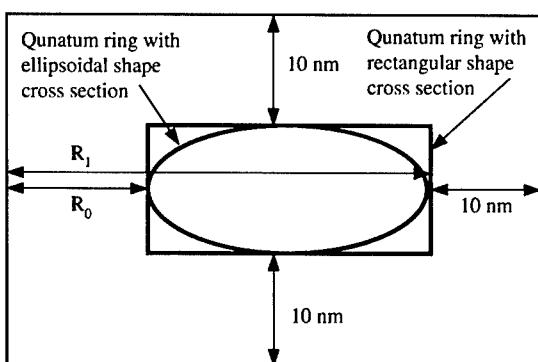


Figure 1. Schematic diagram of cross section view for 3D semiconductor quantum rings.

2. A Quantum Ring Model and Computational Algorithms

We consider electrons confined in quantum ring and use one-band effective Hamiltonian H (Bastard 1988)

$$H = -\frac{\hbar^2}{2} \nabla_r \left(\frac{1}{m(E, r)} \right) \nabla_r + V(r), \quad (1)$$

where $m(E, r)$ is the electron effective mass that depends on energy and position

$$\frac{1}{m(E, r)} = \frac{P^2}{\hbar^2} \left[\frac{2}{E + E_g(r) - E_c(r)} + \frac{1}{E + E_g(r) + \Delta(r) - E_c(r)} \right] \quad (2)$$

and $V(r) = E_c(r)$ is the confinement potential of quantum rings. The $E_c(r)$, $E_g(r)$, $\Delta(r)$, and P are the position dependent electron band edge, band gap, spin-orbit splitting in the valance band, and momentum matrix element, respectively (Bastard 1988). We solve the quantum ring problem with cylindrical coordinate (R, ϕ, z) . The quantum ring system is cylindrical symmetry so that the wave function can be written as: $\Phi(r) = \Phi(R, z) \exp(i\phi)$, where $l = 0, \pm 1, \pm 2, \dots$ is the electron orbital quantum number and the model is written as

$$-\frac{\hbar^2}{2m_i(E)} \left(\frac{\partial^2}{\partial R^2} + \frac{\partial}{R \partial R} + \frac{\partial^2}{\partial z^2} - \frac{l^2}{R^2} \right) \Phi_i(R, z) + V_i(R, z) \Phi_i(R, z) = E \Phi_i(R, z) \quad (3)$$

where $V_{i=1}(R, z) = 0$ is inside the ring and $V_{i=2}(R, z) = V_0$ is outside the ring. The boundary conditions are

$$\begin{aligned} \Phi_1 &= \Phi_2 \quad \text{and} \quad \frac{1}{m_1(E)} \left\{ \frac{\partial \Phi_1}{\partial R} + \frac{df_s}{dR} \frac{\partial \Phi_1}{\partial z} \right\} \\ &= \frac{1}{m_2(E)} \left\{ \frac{\partial \Phi_2}{\partial R} + \frac{df_s}{dR} \frac{\partial \Phi_2}{\partial z} \right\}, \end{aligned} \quad (4)$$

where $z = f_s(R, z)$ is a contour generator of the cross section of quantum rings structure in (R, z) plane. The 3D structures are generated by the rotation of this contour around the z -axis.

The electron effective mass is a spatial and energy dependent function, therefore the derived Schrödinger equation is a nonlinear equation in energy. To obtain a “self-consistent” solution to the model, we propose here a nonlinear iterative algorithm as shown in

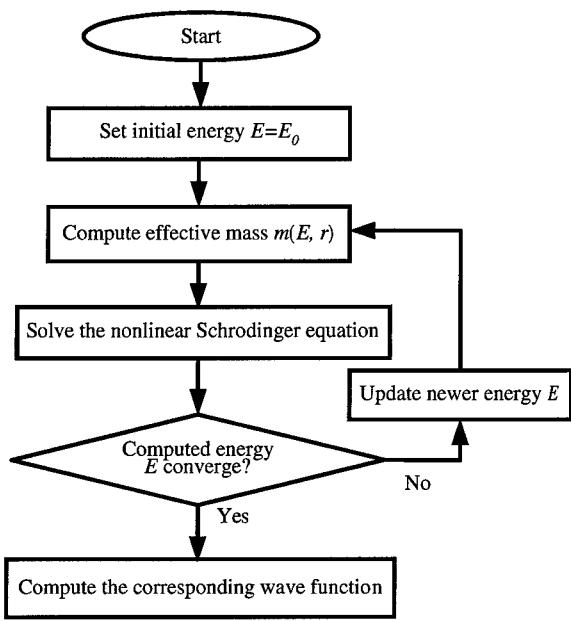


Figure 2. The proposed computational algorithm for semiconductor quantum ring simulation.

Fig. 2. This feedback iteration scheme consists of (i) setting an initial energy E_0 , (ii) computing effective mass m , (iii) solving the Schrödinger equation with its boundary conditions for energy E , and (iv) returning to step (ii). The iteration loop terminates when a

specified stopping criterion is reached. To solve the Schrödinger equation in step (iii), the Schrödinger equation is discretized with nonuniform mesh central difference method, and the corresponding matrix eigenvalue problem is solved with the balanced and shifted QR method, and the inverse iteration method. The dominant method for solving matrix eigenvalue problem in semiconductor nanostructure simulation is the QR algorithm (Watkins 2000, Li *et al.* 2001c). In our simulation experience, convergence (the maximum norm error in energy $< 10^{-8}$ eV) is reached by taking only 9–10 iterations.

3. Calculation Results and Discussions

We focus here the discussions on our simulation results for InAs/GaAs quantum rings with the material parameters. For InAs, the energy gap E_{1g} is 0.42 eV, Δ_1 is 0.38 eV, and $m_1(0) = 0.024m_0$. For GaAs we choose: $E_{2g} = 1.52$ eV, $\Delta_2 = 0.34$ eV, and $m_2(0) = 0.067m_0$. The band offset parameter is taken as $V_0 = 0.77$ eV. The Fig. 3(a) shows the dependence of electron ground state energy on the inner radius (R_0) of InAs/GaAs quantum ring. In this example, we calculated 3D rectangular and ellipsoidal radial cross sections for quantum rings where the height $z_0 = 4$ nm and outer radial width $R_1 = 10$ nm, respectively. The Fig. 3(a) suggests

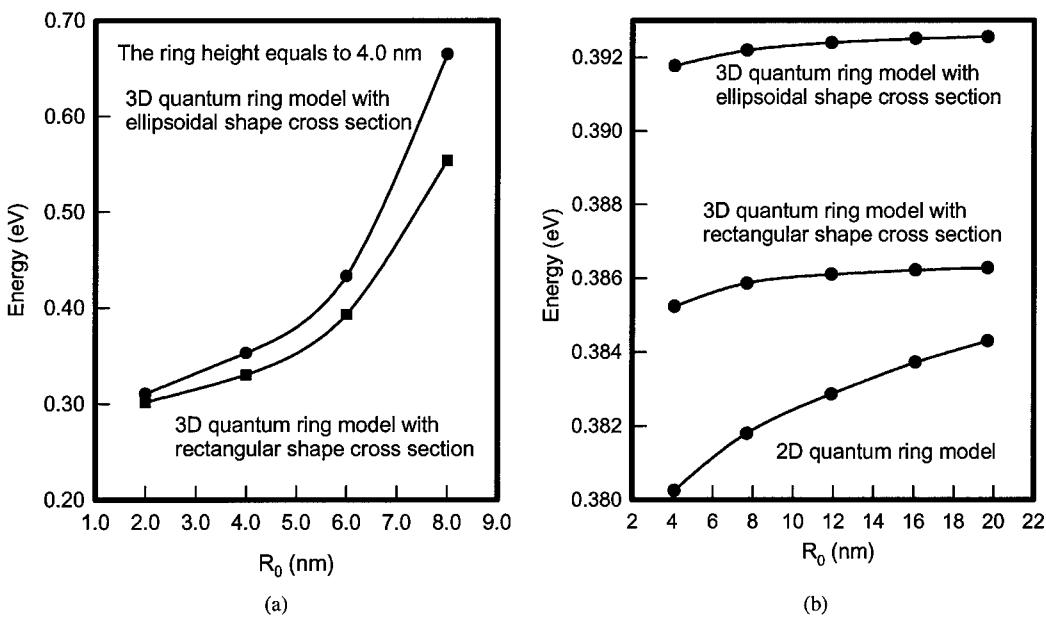


Figure 3. (a) The dependence of electron ground state energy on the inner radius R_0 of InAs/GaAs nano-scopic quantum ring. (b) Electron ground energy states of ring with different physical models.

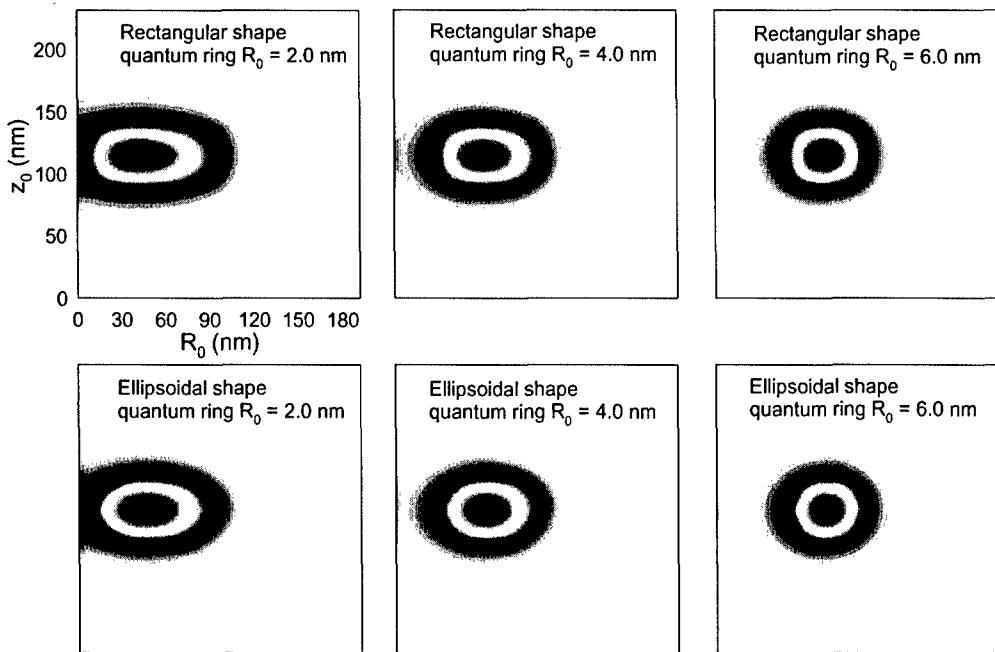


Figure 4. Plots of the localized wave functions transition with various R_0 , where the ring height and outer radius are the same with Fig. 3(a).

that the quantum ring with rectangular shape cross section view (Bruno-Alfonso and Latgé 2000) has a good approximation only when the inner radius is small.

Furthermore, to clarify the model effects for electron ground state energy levels, we also compare 2D simplified (adiabatic) model (Emperador *et al.* 2000, Li *et al.* 2001a) and 3D models with rectangular and ellipsoidal radial cross sections. Figure 3(b) shows the dependence of electron ground state energy on R_0 for ultra thin InAs/GaAs quantum ring. In this simulation, the 2D model and 3D rectangular and ellipsoidal radial cross sections for rings of height $H = 2$ nm and radial width (radius difference) $\Delta R = 20$ nm are computed. We find a large discrepancy among these results and also verify that results for different radial cross shapes (we calculated the rectangular shape (Bruno-Alfonso and Latgé 2000) and ellipsoidal shape as a more realistic (Lorke *et al.* 2000)) are different greatly. In addition, Fig. 4, shows the transition of the localized wave functions versus R_0 .

4. Conclusions

In conclusions, we have studied the electron energy states of realistic 3D nano-scopic semiconductor quantum rings. We treated the problem with the effective one-band Hamiltonian approximation, the position and

energy dependent quasi-particle effective mass approximation, the finite hard wall confinement potential, and the Ben Daniel-Duke boundary conditions simultaneously. The 3D model was solved with the nonlinear iterative algorithm to obtain final self-consistent solutions. This study has presented an alternative to compute the energy levels of nano-scopic semiconductor quantum ring and clarified the principal dependencies of energy states on material band parameter and ring size for various ring shapes. We found that the energy levels strongly depend on the radial cross shapes of InAs/GaAs quantum rings. The dependence of energy states on shapes for 3D nano-scopic quantum ring indicated a significant difference among those reported results derived with simplified 2D approaches.

References

- Bastard, G. 1988. Wave Mechanics Applied to Semiconductor Heterostructures. Les Edition de Physique, Les Ulis.
- Bimberg D. *et al.* 2000. Thin Solid Films 367: 235.
- Bruno-Alfonso A. and Latgé A. 2000. Physical Review B 61: 15887.
- Emperador A. *et al.* 2000. Physical Review B 62: 4573.
- Li Y. *et al.* 2001a. Computer Physics Communications 140: 399.
- Li Y. *et al.* 2001b. Solid State Communications 120: 79.
- Li Y. *et al.* 2001c. Computer Physics Communications 141: 66.
- Lorke A. *et al.* 2000. Physical Review Letter 84: 2223.
- Tsai C.-H. *et al.* 1998. IEEE Photonics Technology Letters 10: 751.
- Watkins D. 2000. Journal of Computational and Applied Mathematics 123: 67.



Fully Numerical Monte Carlo Simulator for Noncubic Symmetry Semiconductors

LOUIS TIRINO, MICHAEL WEBER AND KEVIN F. BRENNAN

School of Electrical and Computer Engineering, 777 Atlantic Dr., Georgia Tech, Atlanta, GA 30332-0250, USA

ENRICO BELLOTTI

Department of Electrical and Computer Engineering, 8 Saint Mary's St., Boston University, Boston, MA 02215-2421, USA

MICHELE GOANO

Dipartimento di Elettronica, Politecnico di Torino, I-10129 Torino, Italy

P. PAUL RUDEN

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA

Abstract. In this paper, we present the workings of a fully numerical Monte Carlo simulator that can be employed to study transport in materials with noncubic symmetry. All of the principal ingredients of the Monte Carlo model, i.e., the energy band structure, phonon scattering rates, and impact ionization transition rate are used in numerical form. Various considerations such as k-space mesh size, numerical integration convergence, etc. that impact numerical accuracy will be discussed. The workings of the simulator are illustrated using example calculations of the bulk transport properties of GaAs and GaN. The simulation of bulk GaAs in particular challenges the numerics since the low electron effective mass within the gamma valley requires a high degree of numerical refinement to correctly capture the dynamics in this region. We calculate the steady-state drift velocity, impact ionization coefficients, valley occupations, and average carrier energy in bulk GaAs and GaN.

Keywords: Monte Carlo, phonon scattering

1. Introduction

In this paper we outline the ingredients of a fully numerical Monte Carlo simulator for the analysis of bulk transport properties in wide band-gap semiconductors. The methods discussed are easily adapted to a fully numerical Monte Carlo device simulator as well. The key numerical ingredients to the simulator are the electronic bandstructure and numerically calculated phonon scattering rates (including acoustic, polar optical, nonpolar optical, and impact ionization transition rates.)

The use of a numerical electronic bandstructure and a quadratic interpolation scheme has been used

successfully and is discussed in detail in Fischetti and Laux (1988) and Fischetti (1991), so it will not be addressed here.

Commonly, however, one of many possible techniques is used to account for and describe carrier scattering within a material, and some of those techniques include analytical scattering expressions for the acoustic, polar and nonpolar optical scattering rates, a total scattering rate, valley assignments and valley-specific scattering, etc.

The problem with these techniques is that they all make, in some way, an approximation and/or assumption based on known behavior(s) of a particular

material. With the emergence of certain semiconductor compounds with few known physical properties, these generalizations are no longer necessarily valid. Therefore, to remove assumptions that may produce acceptable results in one material, but which may be completely or partially wrong in a material with unknown and perhaps more complicated transport properties (e.g. SiC-4H and the elaborate band-crossings), an approach that is more general and comprehensive is needed.

Another aspect of Monte Carlo semiconductor simulation that is critical in ascertaining the correct transport properties is the final state selection after a scattering event.

In this paper, we discuss a technique intended to more accurately capture the most complete description of carrier scattering within the Brillouin Zone for any semiconductor compound. The final state selection mechanism will also be discussed.

2. Calculation of Scattering Rates

A common method for determining the magnitude of a particular scattering mechanism is to formulate the rate using an energy-dependent expression. Typically, one assumes an analytical expression for the band structure and computes the scattering rates accordingly (Jacoboni and Lugli 1989). While good results have been obtained from these analytical, energy-dependent expressions for well studied materials (Canali *et al.* 1975, Shichijo and Hess 1981) several assumptions have been made in their realizations that may not apply to some of the emerging materials.

The numerical scattering rates are calculated according to the following equations for polar optical, nonpolar optical, and acoustic phonon scatterings respectively:

$$\begin{aligned} \frac{1}{\tau_{PO}} &= \left(\frac{2\pi}{a} \right) \left(\frac{q\omega_{op}}{8\pi^2} \right) \left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_{st}} \right) \left[N_{op} + \frac{1}{2} \pm \frac{1}{2} \right] \\ &\times \sum_{Vol,IW} \frac{1}{q^2} I(k, k') \delta[E_f - E_i \pm \hbar\omega_{OP}] \Delta^3 k \\ \frac{1}{\tau_{NP}} &= \left(\frac{\pi(D_t k)^2}{(2\pi)^3 \rho \omega_{NP}} \right) \left[N_{op} + \frac{1}{2} \pm \frac{1}{2} \right] \\ &\times \sum_{Vol,IW} I(k, k') \delta[E_f - E_i \pm \hbar\omega_{NP}] \Delta^3 k \\ \frac{1}{\tau_{AC}} &= \left(\frac{\pi E_1^2}{(2\pi)^3 \rho} \right) \sum_{Vol,IW} \frac{q^2}{\omega(q)} \left[N(q) + \frac{1}{2} \pm \frac{1}{2} \right] \\ &\times I(k, k') \delta[E_f - E_i \pm \hbar\omega(q)] \Delta^3 k \end{aligned}$$

The term of $\pm \hbar\omega$ denotes absorption and emission, respectively.

The term $I(k, k')$ is the squared overlap integral of the Bloch functions of the initial and final states. The final states are chosen from all possible final states in the Brillouin Zone which satisfy the energy conservation requirement. We have selected the polar and nonpolar phonon energies to be constant. For the acoustic phonon scattering rate, ω is calculated as:

$$\omega(q) = \omega_{max} \sqrt{1 - \cos\left(\frac{qa}{4}\right)}$$

where

$$\omega_{max} = \frac{4v_l}{a}$$

for $q < 1.0$, otherwise,

$$\omega(q) = \omega_{max}$$

In the calculation of the scattering rates, there are two important figures of merit that do not appear in the equations. These are the energy delta used in the conservation of energy (the energy range over which the delta function is satisfied) and the resolution of the integration grid within the Brillouin Zone (BZ).

Figure 1 illustrates both the importance of the energy delta as well as the resolution of the integration grid. This figure shows the polar optical scattering rate at Γ ($k = 0, 0, 0$) in GaAs. For coarse integration grids, it is obvious that the rate is strongly dependent on the energy delta. It is clear that since the inclusion of the energy delta arises from numerical considerations and is not a physical parameter, the rate should be independent of this parameter. The dependence is removed when the integration resolution gets finer than $\Delta k = 0.0025$ as can be seen from Fig. 1.

Although the dependence on the energy delta is removed at this point, it can be seen that another important factor in the calculation of the rate is the actual grid resolution as well. Convergence to the value obtained with the analytical expression for $E = 0$ eV, the Γ point in GaAs, is approached as the grid resolution approaches zero.

Figure 2 illustrates the importance of the integration grid resolution. In this figure, the rate is calculated for a collection of query points inside the irreducible wedge (IW). Inspection of Fig. 2 shows that the rate converges for all energy points as the grid resolution becomes

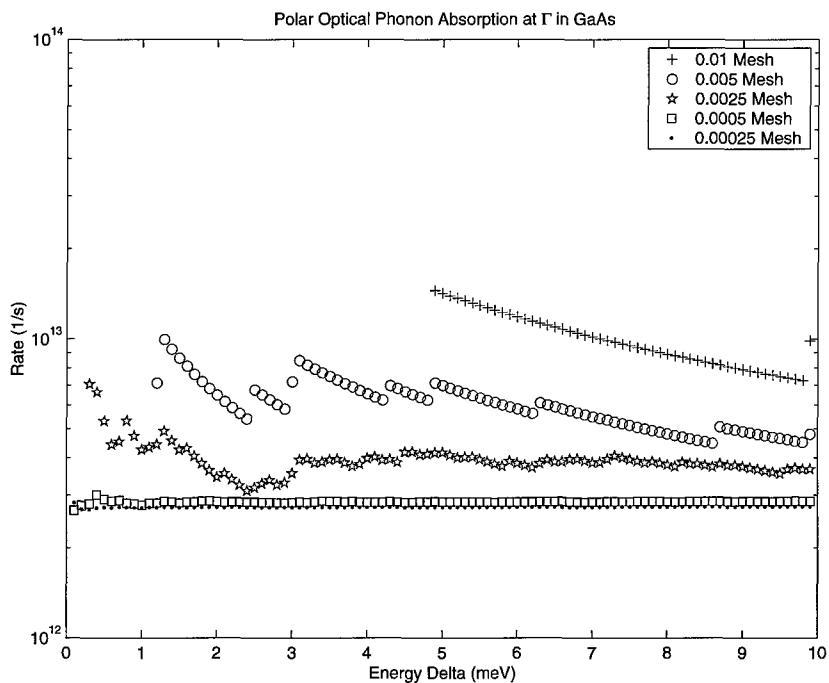


Figure 1. Polar optical phonon absorption at Γ in GaAs. The scattering rate is plotted versus the energy delta for various integration mesh resolutions.

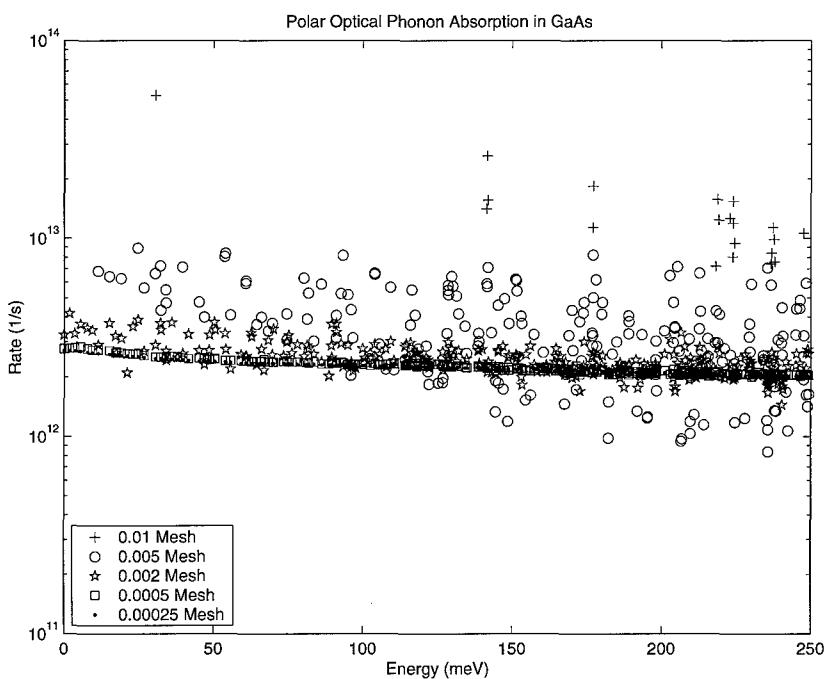


Figure 2. Polar optical phonon absorption for 261 points near Γ in GaAs. The scattering rate is plotted versus the energy of the point.

sufficiently fine. In practice, a grid resolution of 0.0005 is satisfactory, while grid resolutions finer than 0.0005 are computationally impractical.

3. Computational Requirements

It has been noted above that a grid resolution of 0.0005 is sufficient for the calculation of the scattering rates. And while a finer grid would be desirable, it becomes considerably burdensome with respect to the computational power needed; indeed even calculating the rates with a 0.0005 integration grid is intensive and time consuming.

At a grid spacing of 0.0005, there are 668,919,001 points in the IW of a zincblende (ZB) BZ. Our work with GaAs involves 4 conduction bands, so there are therefore effectively over 2.6 billion gridpoints. This exceeds our memory capacity, so the entire grid cannot be generated completely and stored. A scheme is thus needed to integrate over the entire IW. The scheme we employ is to perform the integration in multiple passes. Approximately 60 million points from the wedge are generated in a single pass.

With the parameters described above, there are roughly 45 passes needed to cover the entire IW. There are several possible techniques that can be used to spread the work over multiple processors. The first possible technique that can be used is to spread the query points over different machines. This technique has several flaws. First, every processor will generate the entire IW for a set number of query points. This redundancy of wedge generation amounts to a large waste of computing power. Second, care is needed in judiciously dividing the query points with respect to their energy values. Specifically, if one machine is tasked with calculating the rate for predominantly low energy points, that machine will be generating marginally useful pieces of the IW much of the time, while another machine may be making a large number of calculations much of the time.

A more useful technique is to task different processors to generate only specific pieces of the wedge and calculate the rate for all of the query points in that region. At the end of the calculation a technique for reassembling the different pieces of the total rate across several parts of the IW for a particular point is needed, but it is a trivial exercise. Here, the redundancy is that all of the processors have duplicate copies of the query points, but since the number of query points ($\sim 20,000$)

is far less than the number of points in the IW, the degree of redundancy is less.

Regardless of the technique used in splitting the integration into smaller pieces, a very useful implementation of QuickSort is made in an attempt to reduce the time spent on any given query point. Presume a piece of the IW is generated of size N points. This database of N points consists of the coordinate of the point in the integration grid and its energy. After the points are generated, they are sorted on energy. Then, for a given query point, energy conservation need only be checked for the upper and lower limits of the energy delta being used, with the result being that all points between the upper and lower bounds satisfying energy.

Currently, most of the calculations are being done on Compaq XP1000 workstations with 500 MHz single Alpha processors with either 1GB or 2GB of RAM. This has been sufficient in generating a limited amount of output; a single pass for the polar optical scattering rate for $\sim 30,000$ query points requires approximately 20 hours. In order to generate scattering rates for a sufficient number of materials to perform Monte Carlo bulk and device simulations will necessitate porting the calculations to a larger computational system. This process is currently being implemented.

4. Conclusion

In order to characterize semiconductor materials of unknown properties with a reasonable expectation of accuracy it is necessary to have high confidence in the input ingredients, those being the electronic bandstructure and the scattering rates. The results of the bulk simulations will be presented in later work.

Acknowledgments

This work was sponsored in part by ONR through contract E21-K19, by NSF through grant ECS-9811366 and Yamacraw.

References

- Canali C., Jacoboni C., Nava F., Ottaviani G., and Alberigi-Quaranta A. 1975. Phys. Rev. B 12: 2265.
- Fischetti M.V. 1991. IEEE Trans. Electron. Dev. 38: 634.
- Fischetti M.V. and Laux S.E. 1988. Phys. Rev. B 38: 9721.
- Jacoboni C. and Lugli P. 1989. *The Monte Carlo Method for Semiconductor Device Simulation*. Springer-Verlag, New York.
- Shichijo H. and Hess K. 1981. Phys. Rev. B 23: 4197.



Theoretical Study of RF Breakdown in GaN Wurtzite and Zincblende Phase MESFETs

M. WEBER, L. TIRINO AND K.F. BRENNAN

School of ECE, Georgia Tech, Atlanta, GA 30332-0250, USA

MAZIAR FARAHMAND

Movaz Networks, Old Technology Parkway, South, Norcross, GA 30092, USA

Abstract. In this paper, we present a comparison of the RF breakdown behaviors of representative wurtzite and zincblende phase GaN MESFET structures based on a theoretical analysis. The calculations are made using a full band ensemble, Monte Carlo simulation that includes a numerical formulation of the impact ionization transition rate. Calculations of the RF breakdown voltages are presented for submicron MESFET devices made from either wurtzite or zincblende phase GaN. The devices are otherwise identical. It is found that the RF-breakdown voltage of the devices increases with increasing frequency of the applied large signal RF excitation.

Keywords: Monte Carlo, breakdown, MESFET, high frequency

1. Introduction

The wide band gap semiconductors offer much promise in future high power, high frequency device applications (Eastman 1999, Shur 1998, Trew 1998). Owing to their wide energy band gaps, these materials are less susceptible to high field induced breakdown than conventional silicon or GaAs based devices. Coupled with a higher saturation drift velocity, the high breakdown field strengths of the wide band gap semiconductors offer a significant expansion of the power-frequency coverage range over existing technologies. The higher power density levels that these materials can deliver also provide opportunity for significant miniaturization.

We have examined the device potential of the III-nitride materials using an extension of the materials theory based modeling method (Brennan *et al.* 2000a, b, c, Shichijo and Hess 1981, Kolnik *et al.* 1997, Oguzman *et al.* 1997, Bellotti *et al.* 1999a, b, 2000, Farahmand *et al.* 2001b, Verghese *et al.* 2001, Farahmand and Brennan 1999, 2000) based on a self-consistent, full band Monte Carlo simulation. The device structures investigated were submicron gate length MESFETs.

Both the DC breakdown and frequency performance of these devices were examined.

The materials theory based modeling method has been useful in examining how the transport and device potential of the two polytypes of GaN compare. It has been found that owing to the difference in the band structures that the bulk and device performance of wurtzite and zincblende GaN are substantially different (Oguzman *et al.* 1997, Farahmand and Brennan 2000). Specifically, the DC breakdown voltages of wurtzite and zincblende phase GaN MESFETs have been predicted to be substantially different (Farahmand and Brennan 2000). Knowledge of the breakdown properties of a device is critical in evaluating the maximum output power for a class A amplifier. A class A amplifier has a maximum output power given as,

$$P_{\max} = \frac{(V_{BR} - V_{knee})^2}{8R_L} \quad (1)$$

where V_{BR} is the breakdown voltage, V_{knee} the knee voltage defined as the voltage at which the transistor current saturates and R_L is the load resistance. Obviously the larger the difference in the knee and

breakdown voltages, the greater the maximum output power the device can deliver. The knee voltage is a function of the mobility. A higher mobility results in a smaller value of V_{knee} . The breakdown voltage is a strong function of the energy gap. Wider band gap materials have higher breakdown voltages. Though the energy band gaps of the wurtzite and zincblende phases of GaN are close in magnitude, the breakdown electric field strengths are significantly different. The difference in the breakdown electric field strengths is attributable to the different properties of the wurtzite and zincblende band structures that results in different carrier temperatures within the two materials.

The breakdown voltage has also been experimentally found to depend upon RF conditions. It has been observed that under large-signal high-frequency conditions, devices can be driven beyond their DC breakdown limits (Heo *et al.* 2000, Tkachenko, Wei and Hwang 1996). The breakdown voltage under RF conditions is higher than under DC conditions thus implying that the maximum output power can be greater under RF excitation. Recently, we have examined the RF dependence of the breakdown voltage of zincblende phase GaN MESFETs (Farahmand *et al.* 2001a). It was found that the breakdown voltage under RF drive is frequency dependent however no comparison of the RF breakdown behavior of the different polytypes of GaN was presented. It is the purpose of this paper to present calculated results for both DC and RF breakdown in MESFETs using both the wurtzite and zincblende polytypes of GaN. It is expected that the physical implications derived from the calculations presented herein can impact circuit level designs of high power, high frequency amplifiers relevant to future wireless communications networks.

2. Model Description

The calculations are made using a two-dimensional real space self-consistent, full band ensemble Monte Carlo simulation. The full details of the approach have been presented elsewhere (Farahmand and Brennan 1999) and will not be repeated here. The geometry and doping concentrations of the simulated MESFET device are the same as those reported in Farahmand and Brennan (1999). The small dimensions of the device have been chosen to manage the computational demands of the simulator. Owing to the large number of simulated particles and the relatively long simulation times needed to ensure numerical accuracy, a larger device than that

chosen here is presently unrealistic. The donor doping level of $3 \times 10^{17} \text{ cm}^{-3}$ is typical for GaN devices. All the simulations are performed assuming a constant ambient temperature of 300 K. The dopants are all assumed to be fully ionized and with no doping compensation present.

The device is modeled with two surface depletion regions formed between the source and gate and the drain and gate. It is assumed that the surface states act to deplete out the underlying semiconductor layer resulting in a carrier concentration of 10^{13} cm^{-3} . For simplicity, the depleted region is assumed to be rectangular with a depth equal to half the GaN active layer thickness and a length equal to the source-gate and drain-gate separations.

Aside from the obvious differences in the band structures and the associated phonon and impact ionization rates, the MESFET simulations for the wurtzite and zincblende phases are essentially identical. Treatment of band crossing and mixing points for the wurtzite phase device is performed following the approach outlined in (Farahmand and Brennan 2000).

The breakdown characteristics of the wurtzite and zincblende phase GaN MESFETs are compared under RF operation. A large signal RF bias is applied between the drain and source simulating on-state breakdown. Though in most common source configurations the RF signal is applied to the gate, this situation is more difficult and computationally expensive to simulate using Monte Carlo. The RF breakdown results we present are nevertheless useful in examining the effect on the breakdown voltage of RF excitation since the frequency dependence of the carrier heating is somewhat independent of the bias condition. The waveform applied to the drain contact is assumed to be sinusoidal varying between high and low voltages, V_{hi} and V_{lo} , respectively with angular frequency, ω . The drain current is again calculated under two conditions, with and without impact ionization, in order to determine the breakdown conditions.

The RF breakdown voltage for both the wurtzite and zincblende MESFET structures is determined as follows. In an earlier investigation (Farahmand *et al.* 2001a) we found that for the zincblende phase GaN MESFET that the device is in the breakdown condition defined above (3% difference in the drain current calculated with and without impact ionization present) under a large signal RF voltage described by,

$$V_{ds}(t) = 17.5 + 12.5 \cos(\omega t) \quad (2)$$

with a frequency of 80 GHz. The voltage swings between $V_{hi} = 30$ V and $V_{lo} = 5$ V at a frequency of 80 GHz; the value of V_{hi} is greater than the DC breakdown voltage of the zincblende phase device, 24 V. The frequency of the RF signal is then increased until the device no longer exhibits breakdown. The corresponding frequency is termed the onset breakdown frequency. A similar procedure is used for the wurtzite phase device and the frequency at which the onset of breakdown occurs is compared.

3. RF Breakdown in GaN MESFETs

The RF breakdown dependency is studied by applying a DC bias on the gate, V_{gs} , of -0.1 V with an RF voltage, V_{ds} , described by Eq. (3) applied to the drain with the source grounded. The breakdown behavior is an obvious function of the bias conditions and has been found to also depend upon the frequency of the RF excitation (Farahmand *et al.* 2001a). The earlier calculations made on the zincblende phase MESFET are used as a starting point in these investigations. In

Farahmand *et al.* (2001a) the bias conditions for the zincblende phase MESFET are adjusted such that the device is in breakdown as defined above at a RF frequency of 80 GHz with the applied RF voltage swing given by Eq. (2). In the calculations presented here, we are mainly interested in determining at what RF frequency the device is no longer in breakdown for a given excitation. Owing to the inherent uncertainty of the Monte Carlo generated drain currents, it is best to select a baseline bias condition such that the calculated drain currents with and without impact ionization are significantly different. To this end, we have chosen to use as a starting point a RF frequency of 20 GHz with the applied voltage specified by Eq. (2) for the zincblende device. The applied RF bias in the zincblende device is such that $V_{hi} = 30$ V and $V_{lo} = 5$ V. The resulting calculated drain current is shown in Fig. 1. The solid lines in the figure show the calculated drain current with impact ionization present and the dashed lines show the drain current in the absence of impact ionization. Inspection of Fig. 1(a) shows that at the RF frequency of 20 GHz, the device is well beyond breakdown; the drain currents with and without impact ionization differ by

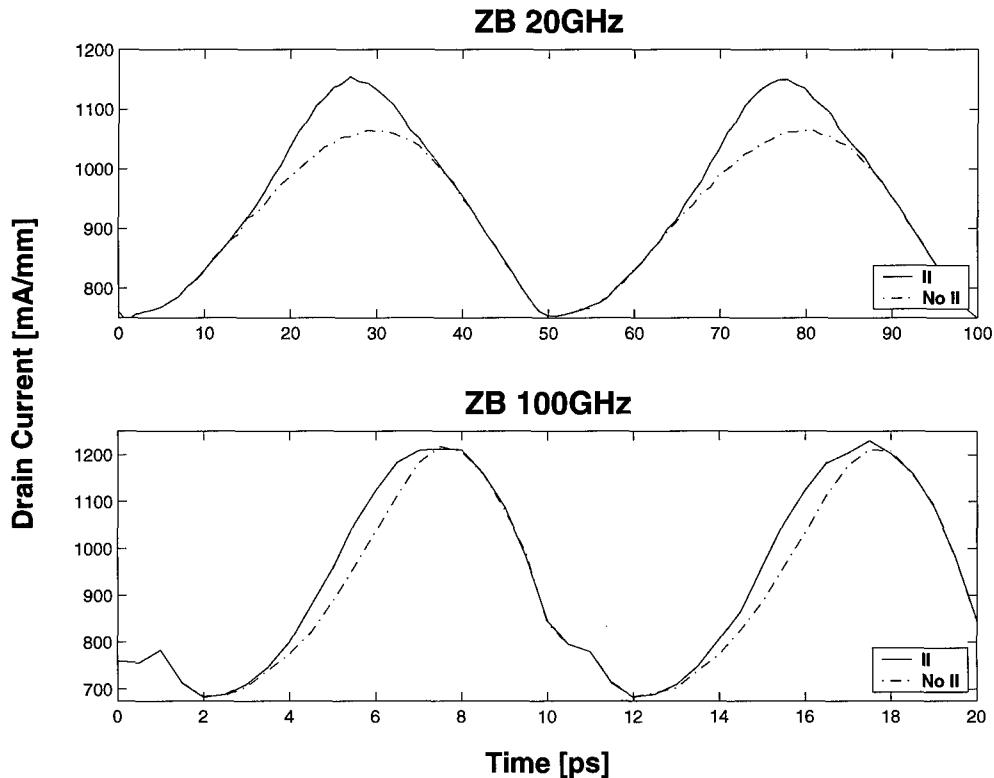


Figure 1. Calculated electron drain current as a function of time with (solid line) and without (dashed line) impact ionization for the zincblende phase GaN MESFET at an RF excitation frequency of (a) 20 GHz and (b) 100 GHz.

about $\sim 8.5\%$, significantly larger than the 3% amount referred to above used to define the breakdown condition. As the frequency is increased the device ultimately is no longer in breakdown. The actual frequency at which the breakdown disappears, defined as the condition where the calculated drain currents with and without impact ionization become the same, is difficult to accurately assess due to the inherent uncertainty in the Monte Carlo calculated currents. Nevertheless, the onset frequency for breakdown lies somewhere in the frequency range between 95–100 GHz for the given bias conditions for the zincblende GaN MESFET. For purposes of illustration, we present the calculated results for a frequency of 100 GHz in Fig. 1(b).

It is interesting to compare the breakdown behavior of the wurtzite and zincblende GaN MESFETs under RF drive. The DC breakdown voltages are different between the wurtzite and zincblende phase devices. The applied gate-source voltage is -0.1 V, the same as for the zincblende phase MESFET. Consequently, the RF voltage applied to the drain must necessarily swing through a higher voltage for the wurtzite phase device

than the zincblende phase device at an RF frequency of 20 GHz. To achieve the same relationship of the calculated drain currents as in the zincblende device, the necessary bias applied to the drain of the wurtzite device is,

$$V_{ds}(t) = 31 + 26 \cos(\omega t) \quad (3)$$

The resulting calculated drain currents with and without impact ionization for the wurtzite phase device are shown in Fig. 2(a) at a RF frequency of 20 GHz. Comparison of Figs. 2(a) and 1(a) show that the ZB and WZ phase devices operate at about the same breakdown point under these conditions. As the frequency is increased breakdown within the wurtzite device also disappears. In this case, the frequency range at which the device is no longer in breakdown is calculated to be 55–60 GHz. For illustration purposes, the calculated drain currents for an RF excitation frequency of 60 GHz are illustrated in Fig. 2(b).

Interestingly, the frequency at which the device no longer exhibits breakdown is lower for the wurtzite

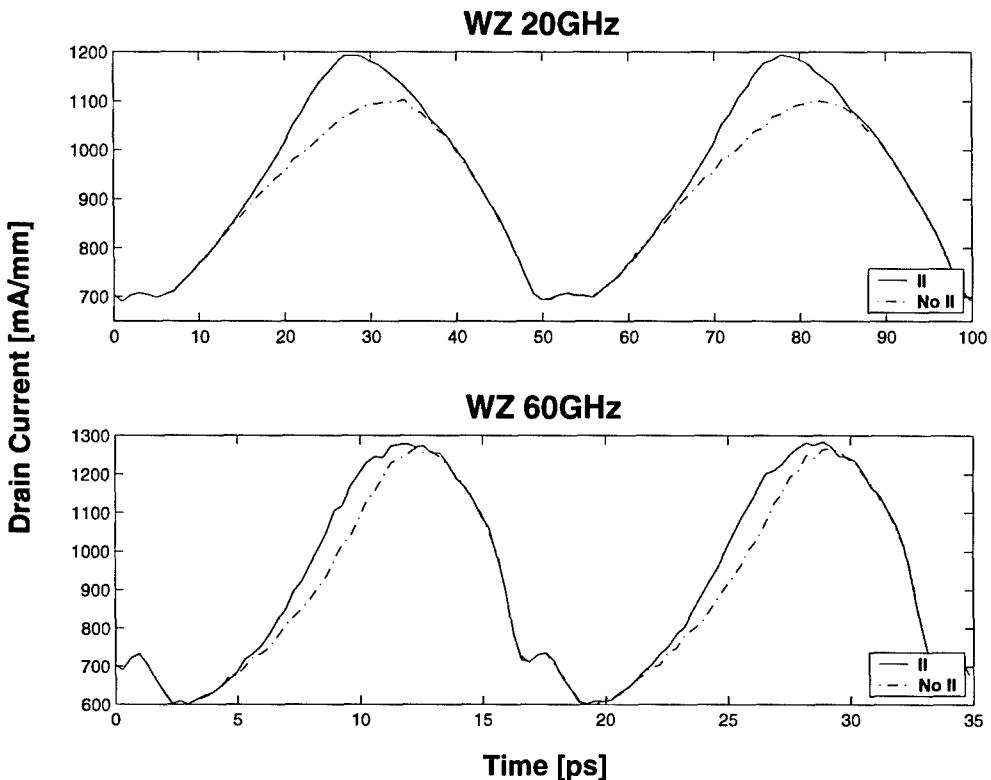


Figure 2. Calculated electron drain current as a function of time with (solid line) and without (dashed line) impact ionization for the wurtzite phase GaN MESFET at an RF excitation frequency of (a) 20 GHz and (b) 60 GHz.

phase device than the zincblende phase device. A possible explanation of this observation is given as follows. In our earlier investigation of RF breakdown (Farahmand *et al.* 2001a) it was suggested that as the RF frequency of the excitation increases, that the electrons can no longer fully respond to the changing electric field. As a result, their energy and consequently their ionization coefficient approach an intermediate value between the two extremes produced by the high and low field components of the RF signal. Thus at some frequency the carriers experience an average field strength that, depending upon the RF signal magnitude, is below that needed for breakdown. Here we have further determined that there is a difference in the frequency for which the device is no longer in breakdown for WZ and ZB phase GaN devices. It is found that the frequency is lower in the WZ phase device than the ZB device when the device leaves breakdown. We speculate that the physical explanation of this effect is that the electrons in WZ GaN cannot follow the signal as rapidly as electrons in ZB GaN. We believe that this is again due to the difference in the density of states between the two phases resulting in a somewhat greater electron inertia in WZ than in ZB. It should be noted that a definitive explanation is presently lacking pending further investigations of the breakdown behavior of devices made using other materials systems. Such a study will be reported in a future work.

4. Conclusion

In this paper, we have presented ensemble, full band, self-consistent Monte Carlo calculations of the RF breakdown behavior of otherwise identical zincblende and wurtzite phase GaN MESFETs.

Investigation of the RF breakdown behavior of wurtzite and zincblende phase GaN MESFETs shows that the breakdown voltage in both materials is highly sensitive to the RF excitation frequency. As the frequency increases, the RF breakdown voltage increases. As the RF frequency increases, the electrons can no longer fully respond to the changing electric field. As a result, the electron energy and consequently the impact ionization coefficient are lowered thereby increasing the breakdown voltage of the device. It is further found that a higher frequency change is required in the ZB phase than the WZ phase to eliminate breakdown in the corresponding MESFET structures.

Acknowledgments

This work was sponsored in part by the National Science Foundation through Grant ECS-9811366, the Office of Naval Research through Contract E21-K19, and the Yamacraw Initiative.

References

- Ambacher O., Foutz B., Smart J., Shealy J.R., Weimann N.G., Chu K., Murphy M., Sierakowski A.J., Schaff W.J., Eastman L.F., Dimitrov R., Mitchell A., and Stutzmann M. 2000. *J. Appl. Phys.* 87: 334.
- Bellotti E., Doshi B.K., Brennan K.F., Albrecht J.D., and Ruden P.P. 1999a. *J. Appl. Phys.* 85: 916.
- Bellotti E., Nilsson H.-E., Brennan K.F., and Ruden P.P. 1999b. *J. Appl. Phys.* 85: 3211.
- Bellotti E., Nilsson H.-E., Brennan K.F., Ruden P.P., and Trew R. 2000. *J. Appl. Phys.* 87: 3864.
- Brennan K.F., Bellotti E., Farahmand M., Haralson II J., Ruden P.P., Albrecht J.D., and Sutandi A. 2000a. *Solid-State Electron* 44: 195.
- Brennan K.F., Bellotti E., Farahmand M., Nilsson H.-E., Ruden P.P., and Zhang Y. 2000b. *IEEE Trans. Electron Dev.* 47: 1882.
- Brennan K.F., Kolnik J., Oguzman I.H., Bellotti E., Farahmand M., Ruden P.P., Wang R., and Albrecht J.D. 2000c. In: Pearton S.J. (Ed.), *GaN and Related Materials II*, Vol. 7. Gordon and Breach, Australia.
- Eastman L.F. 1999. *Phys. Stat. Sol. (a)* 176: 175.
- Farahmand M. and Brennan K.F. 1999. *IEEE Trans. Electron Dev.* 46: 1319.
- Farahmand M. and Brennan K.F. 2000. *IEEE Trans. Electron Dev.* 47: 493.
- Farahmand M., Brennan K.F., Gebara E., Heo D., Suh Y., and Laskar J. 2001a. *IEEE Trans. Electron Dev.* 48: 1844.
- Farahmand M., Garetto C., Bellotti E., Brennan K.F., Goano M., Ghillino E., Ghione G., Albrecht J.D., and Ruden P.P. 2001b. *IEEE Trans. Electron Dev.* 48: 535.
- Heo D., Yoo S., Chen E., Gebara E., Hamai M., and Laskar J. 2000. In: Proc. IEEE Int. Microw. Symp. Boston, MA.
- Kolnik J., Oguzman I.H., Brennan K.F., Wang R., and Ruden P.P. 1997. *J. Appl. Phys.* 81: 726.
- Oguzman I.H., Bellotti E., Brennan K.F., Kolnik J., Wang R., and Ruden P.P. 1997. *J. Appl. Phys.* 81: 7827.
- Shichijo H. and Hess K. 1981. *Phys. Rev. B* 23: 4197.
- Shur M.S. 1998. *Solid-State Electron* 42: 2131.
- Smith D.L. 1986. *Solid-State Commun.* 57: 919.
- Tkachenko Y.A., Wei C.J., and Hwang J.C.M. 1996. In: Proc. 47th ARFTG Conf. Dig. San Francisco, CA, p. 67.
- Trew R.J. 1998. In: Willardson R.K. and Beer A.C. (Eds.), *Semiconductors and Semimetals*, Vol. 52. Academic Press, New York, p. 237.
- Vergheese S., McIntosh K.A., Molnar R.J., Mahoney L.J., Aggarwal R.L., Geis M.W., Molvar K.M., Duerr E.K., and Melngailis I. 2001. *IEEE Trans. Electron Dev.* 48: 502.



Quantum Mechanical Model of Electronic Stopping Power for Ions in a Free Electron Gas

YANG CHEN,* DI LI, GENG WANG,[†] LI LIN, STIMIT OAK, GAURAV SHRIVASTAV,
AL F. TASCH AND SANJAY K. BANERJEE

University of Texas at Austin, Austin, TX 78758, USA

Abstract. The electronic stopping power of a free electron gas on a moving charged particle (ion) in a solid is analyzed in the coordinate system moving with the charged particle. By quantum mechanically treating the momentum transfer between the charged particle and the electron gas, explicit analytic expressions for electronic stopping have been derived for ions of all energies in the nonrelativistic regime. The explicit result reduces to well-known results at both high and low ion energies.

Keywords: ion implant model, electronic stopping power

Introduction

The interaction between an ion and an electron gas has been of interest for decades for modeling ion implantation. Comprehensive reviews can be found in Sigmund (1998) and Ziegler, Bieresack and Littmark (1985). For the electronic stopping power of a charged particle in a solid, previous physical pictures based on linear response theory approached the problem by attempting to expand the ion's electric field into a number of $E(k, \omega)$, and integrating the effect of each $E(k, \omega)$, where $E(k, \omega)$ is the Fourier component of the varying electric field in space and time. Based on this picture, the stopping power of matter for a charged moving particle is, in general, expressed as an integral over k and ω (Lindhard 1954, Lindhard and Wihther 1964). The integral is well known as the *dielectric expression*. Later, another study within the dielectric scheme obtained the charge state of swift ions in a solid (Brandt and Kitagawa 1982). In the high-energy regime all the way to relativistic, the Bethe-Bloch (Bethe 1930, 1932, Bloch 1933) formula prevails. The difference between the Bethe-Bloch and linear dielectric approach is that

linear dielectric theory considers the electronic stopping as a *mean energy loss* to the electron gas over all electron transitions, while Bethe-Bloch formula treats electronic stopping in terms of *momentum transfer*. In this paper, we consider electronic stopping as mean momentum transfer to the electron gas at all energies, and treat the momentum transfer quantum mechanically. The subtle conceptual distinction leads to an explicit solution of the electronic stopping power for particles with screened Coulomb potentials.

Description of the Physical Picture

First, we propose a physical picture of electronic stopping power for a moving ion by a free electron gas. The picture is based on quantum mechanics and the concept of momentum transfer. In the laboratory frame of reference, the ion is moving while the electrons in the solid occupy a Fermi sphere centered at the origin. However, in the coordinate system moving with the ion, the ion becomes stationary, and the electrons in the solid occupy a shifted Fermi sphere in reciprocal space. The ion scatters electrons to empty states in the displaced Fermi sea, that have the same energy as the original electrons in the moving coordinates. Since each scattering changes the electron's momentum, the momentum change acts like a force on each electron.

*Present address: Advanced Micro Devices, Sunnyvale, CA 94086, USA.

[†]Present address: IBM, East Fishkill, NY, USA.

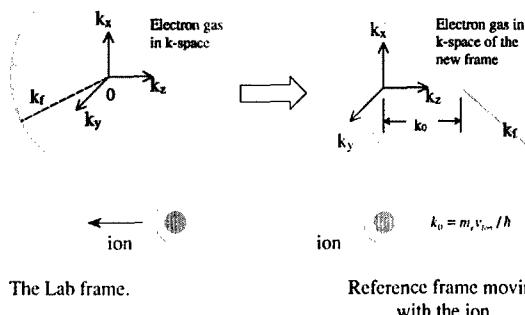


Figure 1. In the k -space of the coordinate system moving with the ion, the ion is stationary, and the electrons occupy a displaced Fermi sphere. The electron Fermi sphere's radius is k_F .

Due to momentum conservation, an equal amount of force is exerted on the ion. The summation of these forces amounts to the total electronic stopping force on a charged particle. Using well-known results of Fermi golden rule, explicit analytic results have been obtained for all ion speeds in the nonrelativistic regime.

Figure 1 shows the distribution of occupied electron states in k -space of the coordinate system moving with the particle. As shown in the figure, the Fermi sphere is displaced from its origin. In different moving coordinates, the electron distribution is a different displaced Fermi sphere. Although this is not an equilibrium state in the moving frame, it is maintained as long as there exist no external perturbations. When a stationary ion is present, the particle serves as a constant perturbation. This perturbation will scatter the electrons to states at the same energy because of the asymmetric distribution of the electron gas in the k -space. The charged particle scatters the electrons in the Fermi sphere to all other states that have the same energy range as the displaced Fermi sphere. The final states region is a larger sphere excluding the displaced Fermi sphere. When the ion's energy is zero, the two spheres overlap with each other so there will be no stopping force. Viewing this picture back in the lab frame, the ion scatters the electron to empty states distributed in the forward direction of the ion.

Each time an electron is scattered from a state k to a new state k' , a momentum transfer of $p' - p = \hbar(k' - k)$ occurs to the electron. On average, the electron momentum transfer times the transition rate of all states amounts to a force on the electron. The charged particle experiences the same amount of force by the electron

in the opposite direction. The contribution due to scattering from all occupied states to all allowed states accounts for the total electronic stopping power of the charged particle.

Analytical Solution

In the ion's coordinate system, consider a single free electron with a wave function $|\psi\rangle = \frac{e^{i\vec{p}\cdot\vec{r}}}{\sqrt{V}}$. Its average momentum is $\langle \vec{p}(t) \rangle = \langle \psi(t) | \vec{p} | \psi(t) \rangle$

In the presence of an ion as a perturbation, the momentum changes with time. This, in effect, can be viewed as a force exerted on the electron:

$$\frac{dp(t)}{dt} = \frac{d}{dt} \langle \psi(t) | \vec{p} | \psi(t) \rangle \quad (1)$$

If one considers the ion as a perturbation, the perturbed wave function can be written as follows:

$$|\psi(t)\rangle = \sqrt{1 - \sum_{k'} |c_{k'}(t)|^2} |\psi_k\rangle + \sum_{k'} c_{k'}(t) |\psi_{k'}\rangle \quad (2)$$

It should be noted although many perturbation treatments approximate the square root term as unity (Schiff 1968), it is important not make that approximation here. Then,

$$\begin{aligned} p(t) &= \langle \psi(t) | \vec{p} | \psi(t) \rangle = \left(1 - \sum_{k'} |c_{k'}(t)|^2\right) \langle \psi_k | p | \psi_k \rangle \\ &\quad + \sum_{k'} |c_{k'}(t)|^2 \langle \psi_{k'} | p | \psi_{k'} \rangle \\ &= p + \sum_{k'} |c_{k'}(t)|^2 (\hbar k' - \hbar k) \end{aligned}$$

Therefore

$$\frac{d}{dt} p(t) = \sum_{k'} \frac{d}{dt} |c_{k'}(t)|^2 (\hbar k' - \hbar k) \quad (3)$$

The physical meaning of this equation is clear: the momentum change of the electron gas is the transition rate times the momentum difference between original and new states.

Quantum mechanically (Schiff 1968),

$$\frac{d}{dt} |c_{k'}(t)|^2 = w_{kk'} = \frac{2\pi}{\hbar} |H_{kk'}|^2 \delta(E_k - E_{k'}) \quad (4)$$

where $H_{kk'}$ is the matrix element of the Hamiltonian.

Also one can let

$$\sum_{k'} \rightarrow \frac{V}{8\pi^3} \int_k dk'^3 = \frac{V}{8\pi^3} \int d\Omega \int_0^\infty k'^2 dk' \quad (5)$$

Thus

$$\begin{aligned} \frac{d}{dt} \vec{p}(t) &= \frac{V}{8\pi^3} \int d\Omega \int_0^\infty k'^2 \delta(E_k - E_{k'}) dk' \\ &\times (\hbar \vec{k}' - \hbar \vec{k}) \frac{2\pi}{\hbar} |H_{kk'}|^2 \end{aligned} \quad (6)$$

For a screened Coulomb potential with screening wave number λ

$$V(r) = \frac{Ze^2}{r} \exp(-\lambda r) \quad (7)$$

which can represent many types of ions and charged particles. Then (Schiff 1968),

$$H_{kk'} = \frac{1}{V} \frac{4\pi Ze^2}{\lambda^2 + (\vec{k} - \vec{k}')^2} \quad (8)$$

Plugging Eq. (8) into Eq. (6) and integrating, one obtains:

$$\begin{aligned} \frac{d}{dt} p(t) &= 8\pi \frac{(Ze^2)^2}{V} \frac{m}{\hbar^2} \left(\frac{1}{4k^2} \ln \left(\frac{4k^2 + \lambda^2}{\lambda^2} \right) \right. \\ &\left. - \frac{1}{(4k^2 + \lambda^2)} \right) \end{aligned} \quad (9)$$

This equation represents the stopping power of a single electron on a moving ion with screened Coulomb potential. The direction of the force is in the electron's direction of momentum in the ion's frame, directly opposite to the ion's motion.

In order to calculate the stopping force of the entire ensemble of electrons, one should calculate the momentum transfer rate due to scattering from all occupied states to all allowed empty states. The single electron stopping force Eq. (9) assumes all other states are allowed. However, for an electron gas, electrons will not be scattered to those states already occupied in the Fermi sphere. Therefore, for each electron, one should exclude the stopping power contributions due to the scattering to occupied states. Nevertheless, the above expression for the single electron stopping force Eq. (9) will be used for the following reasons.

As shown in Fig. 2, the contribution due to scattering from state k to k' , and the contribution due to scattering from state k' to k both need be excluded from the total force. The contribution of $dk^3 \rightarrow dk'^3$ is: $(k' - k)w_{kk'}$, while the contribution of $dk'^3 \rightarrow dk^3$ is $(k - k')w_{k'k}$.

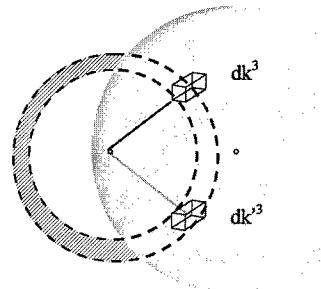


Figure 2. For the scattering between dk^3 and dk'^3 within the sphere, the contribution due to $dk^3 \rightarrow dk'^3$ cancels with the contribution due to $dk'^3 \rightarrow dk^3$. Therefore, the results will not be affected by including these two contributions. This substantially facilitates the calculation of the total stopping power.

Since $w_{kk'} = w_{k'k}$, the two contributions cancel each other. Therefore, the total force can be considered to be simply Eq. (9) integrated over the entire Fermi sphere.

The integration is performed separately for two cases, the case of $v_{ion} < v_F$, for which the $k = 0$ origin is in the shifted Fermi-sphere; and the case of $v_{ion} > v_F$, for which the center is outside the shifted Fermi-sphere. V_F is the Fermi velocity of the electron gas.

For the case of $v_{ion} > v_F$, the integration is performed over the entire Fermi sphere.

$$\begin{aligned} F_{ion} &= \sum_{k \in \text{fermi sphere}} F(k) = \frac{V}{8\pi^3} g \int_{k \in \text{fermi sphere}} F(k) dk^3 \\ &= \frac{V}{8\pi^3} g \int_{k \in \text{fermi sphere}} F(k) \frac{\vec{k}}{|k|} dk^3 \end{aligned} \quad (10)$$

where $g = 2$ is the spin degeneracy of electron.

The integration yields for the stopping power:

$$\begin{aligned} \frac{dE}{dx} &= g \left(\frac{1}{8\pi^3} \right)^2 \frac{(\pi)(\pi Ze^2)^2 m}{k_0^2 \hbar^2} (2\pi)^2 \frac{1}{72\lambda} \\ &\times \left\{ -4k_F \lambda (156k_0^2 + 196k_F^2 + 15\lambda^2) \right. \\ &+ (144(k_0^2 - k_F^2)^2 - 216(k_0^2 + k_F^2)\lambda^2 - 15\lambda^4) \\ &\times \left(\text{ArcTan} \left(\frac{2(k_0 - k_F)}{\lambda} \right) \right. \\ &- \left. \text{ArcTan} \left(\frac{2(k_0 + k_F)}{\lambda} \right) \right) + 192\lambda \\ &\times \left[(k_F^3 - k_0^3) \ln \left(1 + \frac{4(k_0 - k_F)^2}{\lambda^2} \right) \right. \\ &\left. + (k_F^3 + k_0^3) \ln \left(1 + \frac{4(k_0 + k_F)^2}{\lambda^2} \right) \right] \end{aligned} \quad (11)$$

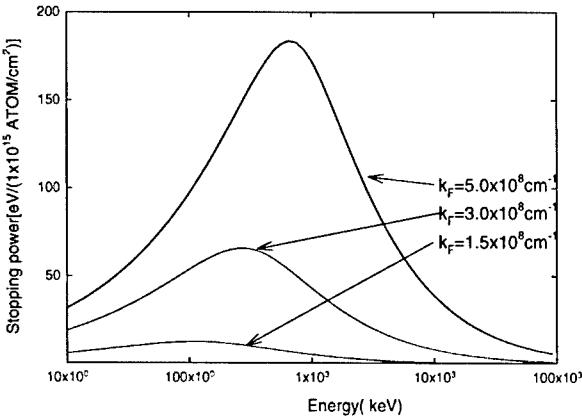


Figure 3. Plot of electronic stopping power over the whole energy range for an ion with $\lambda = 2.0 \times 10^8 \text{ cm}^{-1}$ in three different solids with different k_F .

where k_0 is the average displacement, $\hbar k_0 = m_e v_{ion}$, and k_F is Fermi wave vector, $\frac{8\pi}{3} k_F^3 = n_0$, where n_0 is the electron density.

For $v_{ion} < v_F$, the integration yields:

$$\begin{aligned} \frac{dE}{dx} = & \left(\frac{1}{8\pi^3} \right)^2 \frac{(\pi g(\pi Ze^2)^2 m)}{k_0^2} \frac{1}{\hbar^2 (2\pi)^2} \frac{1}{72\lambda} \\ & \times \left\{ -4k_0\lambda (156k_F^2 + 196k_0^2 + 15\lambda^2) \right. \\ & + (144(k_0^2 - k_F^2)^2 - 216(k_0^2 + k_F^2)\lambda^2 - 15\lambda^4) \\ & \times \left(\text{ArcTan}\left(\frac{2(k_F - k_0)}{\lambda}\right) \right. \\ & - \text{ArcTan}\left(\frac{2(k_0 + k_F)}{\lambda}\right) \left. \right) + 192\lambda \\ & \times \left[(k_0^3 - k_F^3) \ln\left(1 + \frac{4(k_0 - k_F)^2}{\lambda^2}\right) \right. \\ & \left. \left. + (k_F^3 + k_0^3) \ln\left(1 + \frac{4(k_0 + k_F)^2}{\lambda^2}\right) \right] \right\} \quad (12) \end{aligned}$$

The above results should be compared with the dielectric expression of Lindhard (1954) and Lindhard and Wihther (1964) and the result of Brandt and Kitagawa (1982). In order to facilitate the comparison with these previous results, Eqs. (11) and (12) can be named *scattering expressions*. A plot of the electronic stopping power over the whole range is shown in Fig. 3.

Discussion

Ziegler, Biersack and Littmark (1985) has considered the stopping power of heavy ions separately in the “low

velocity” ($v_0 < v_F$) and “high velocity” ($v_0 > 3 \times v_F$) regimes. The Fermi velocity has been first been used as a separation point by Brandt and Kitagawa (1982). As shown above, this study supports the fact that the Fermi velocity plays an important role in the separation of low and high energies. However, the Fermi velocity in this calculation appears more naturally depending on whether the shifted Fermi sphere includes the origin, while other theories attribute the distinction to the difference in the charged state of the ion (Brandt and Kitagawa 1982).

For a low velocity ($v < v_F$) heavy ion, the Taylor expansion of the electronic stopping power near $k_0 = 0$ yields:

$$\begin{aligned} \frac{dE}{dx} = & \frac{4(Z)^2}{3\pi} E_0 \frac{\left(-4k_F^2 + (4k_F^2 + \lambda^2) \ln\left(1 + \frac{4k_F^2}{\lambda^2}\right) \right)}{(4k_F^2 + \lambda^2)} \\ & \times k_0 + O(k_0^3) \quad (13) \end{aligned}$$

which means $F(v) \propto k_0 \propto v$ for low ion velocities. It is also seen that the k_0^2 term in Eq. (13) is zero. This means the $F(v) \propto v$ property is maintained for a rather large range of ion velocities below v_F . This is supported by many experimental studies and empirical models (Ziegler, Biersack and Littmark 1985).

For $v_{ion} \ll v_F$, Bonderup (Xia and Tan 1986) obtained a result identical to Eq. (13) also by considering the electron transition at the Fermi surface in the frame moving with the ion. The displaced Fermi sphere was considered as causing a “net current” to the ion, and the current was calculated by considering the Boltzmann relaxation time of the electron gas.

With increasing energy, the stopping power increases until a maximum is attained. After that, the stopping power decreases due to the decreased scattering rate. At high energies, the electron’s Fermi sphere moves further away from the origin. It looks as if the whole Fermi sphere acts as a single entity consisting of the entire ensemble of electrons. Therefore, if we do not consider relativistic velocities, the electronic stopping power can be simplified as:

$$\begin{aligned} \frac{dE}{dx} = & F_{electron} N = n_0 \frac{1}{8\pi^3} \frac{(4\pi Ze^2)^2 m}{1} \frac{1}{\hbar^2 (2\pi)^2} \\ & \times \left(\frac{1}{4k^2} \ln\left(\frac{4k^2 + \lambda^2}{\lambda^2}\right) - \frac{1}{(4k^2 + \lambda^2)} \right) \\ \cong & n_0 (Ze^2)^2 \frac{\pi}{E} \ln\left(\frac{4k^2}{\lambda^2}\right) \quad (14) \end{aligned}$$

which reduces to the non-relativistic version of Bethe-Bloch formula (Bethe 1930, 1932, Bloch 1933). A

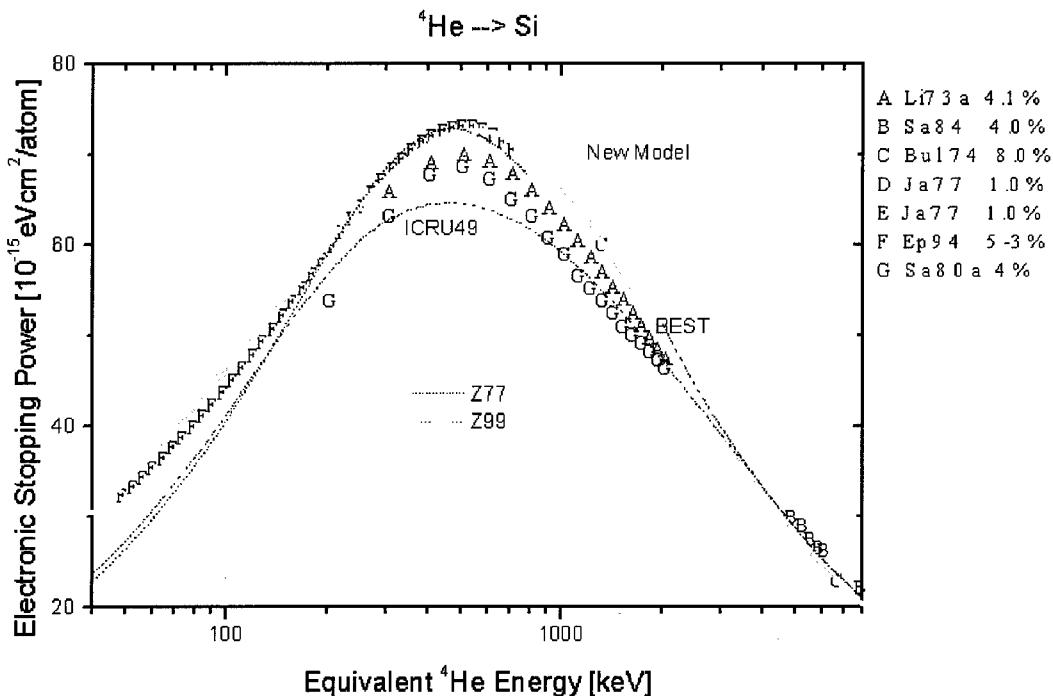


Figure 4. Comparison of the new model with a number of experimental data and empirical models Paul, <http://www.exphys.uni-linz.ac.at/stopping/> for He+ implant into Si. The thin lines and discrete characters are various empirical models and experimental data (Paul op cit). The thick line is the new model results.

precise relativistic calculation needs to consider the distortion of the Fermi sphere due to the Lorentz transformation. It is concluded that the reduction of electronic stopping power at high energies is due to the reduced scattering rate at high energies.

Figure 4 compares the new model with previous experimental data and empirical models. As shown in the figure, the model performs well over a whole range of energy.

Conclusions

In conclusion, the paper provides a new perspective based purely on quantum mechanics to consider electronic stopping of a charged particle in a solid. In previous theories, different stopping mechanisms are proposed and work in different energy ranges. The present paper attempts to attribute electronic stopping to a single mechanism that works for all non-relativistic energies. Scattering-induced momentum transfer between the ion and the electron gas should be considered as the primary part of the electronic stopping power.

The paper gives a theoretical calculation of electronic stopping power for all energies of a screened charged particle. The explicit result reduces to well

known results at both high and low ion energies. It is concluded that this quantum mechanical model will be sufficient to predict most of the electronic stopping behavior in the non-relativistic regime.

Acknowledgment

This work is supported in part by Semiconductor Research Corporation.

References

- Bethe H. 1930. Ann. Physik 5: 324.
- Bethe H. 1932. Z. Physik 76: 293.
- Bloch F. 1933. Ann. Physik 16: 285.
- Brandt W. and Kitagawa M. 1982. Phys. Rev. B 25(9): 5631–5637.
- Lindhard J. 1954. Mat. Fys. Medd. 28(8).
- Lindhard J. and Wihther A. 1964. Mat. Fys. Medd. 34(4).
- Paul H. Stopping power for light ions. Available at <http://www.exphys.uni-linz.ac.at/stopping/>.
- Schiff L. 1968. Quantum Mechanics. McGraw-Hill, New York.
- Sigmund P. 1998. Stopping power in perspective. Nucl. Inst. & Meth. B 135(1–3): 1–15.
- Xia Y. and Tan C. 1986. Nucl. Inst. & Meth. B B13(1–3): 100–106.
- Ziegler J.F., Biersack J.P., and Littmark U. 1985. The Stopping and Range of Ions in Solid. Pergamon Press.



An Analytical 1-D Model for Ion Implantation of Any Species into Single-Crystal Silicon Based on Legendre Polynomials

G. SHRIVASTAV, D. LI, Y. CHEN, G. WANG, L. LIN, S. OAK, A.F. TASCH AND S.K. BANERJEE
University of Texas at Austin, Austin, TX 78758, USA

Abstract. A computationally-efficient 1-D analytical model for ion implantation of any species into single crystal silicon is presented. By interpolating between a few species, the model can predict as-implanted profiles for all the other species. The model uses Legendre polynomials as basis functions. The results of the model are in good agreement with UT-MARLOWE, which is a physically-based and experimentally verified Monte Carlo simulator.

Introduction

As the device geometries shrink to the deep-submicron regime, ion implantation continues to be an important part of the fabrication process. Accurate and computationally-efficient ion implantation models that properly account for various implant parameters are therefore needed for both CMOS design and modeling of processes such as annealing. Further, more and more new species are being implanted into Si. It is therefore necessary to develop a model, which accurately models ion implantation for any arbitrary species into Si.

We previously developed a physically-based “Universal” model for this purpose, using the Monte Carlo approach, for our ion implant simulator UT-MARLOWE (Tasch, Yang and Morris 1995, Morris *et al.* 1996). This model has been verified extensively by comparing the resulting profiles with Secondary Ion Mass Spectrometry (SIMS) profiles (Chen 2001). However the Monte Carlo approach requires rather long computational time since it computes the trajectory of each simulated ion after every collision.

Further, if one wants to reduce the statistical noise inherent in the Monte-Carlo process, one may have to simulate the implantation of a very large number of ions. Analytical models, on the other hand, require much less computation time and resources. We recently developed an analytical model for ion implantation that uses Legendre polynomials as its basis (Balamurugan *et al.* 1998). This model uses parameterized functions to model the as-implanted profile and hence is much

more computationally efficient than the Monte Carlo based UT-MARLOWE.

Any new species can be simulated using this model by generating a set of parameters that characterize the dependence of the implant profile for this species on various implant parameters, such as energy, dose, tilt and rotation angles and the thickness of the screen oxide, nitride or silicide layer. However, generating tables for every species that is to be simulated is time taking and cumbersome. A model that could predict the as implanted profile for a new species by taking advantage of the existing tables for other species and using some kind of interpolation would therefore be very useful. A 1-D analytical model using this approach is presented in this paper.

Analytical Model

The model uses Legendre polynomials as a fitting function for the impurity profiles. The Legendre polynomials have an advantage over the widely used dual-Pearson model since, unlike the latter, these form a completely orthonormal basis and hence the parameter extraction procedure is not dependent on the initial guess. The parameter extraction procedure is thus greatly simplified.

The impurity profiles are modeled as a linear combination of the first 14 Legendre polynomials.

$$C(z) = \sum_0^{13} a_i L_i(z') \quad (1)$$

and

$$z' = b_1 z + b_2 \quad (2)$$

where $C(z)$ is the impurity concentration at depth z , $L_i(z')$ is the Legendre polynomial of degree i , b_1 and b_2 are the constants that map z from the interval $[z_{\min}, z_{\max}]$ to $[-1, 1]$ and z_{\min} and z_{\max} are the depths between which the impurity concentration lies within the range of interest.

Since the Legendre polynomials are orthonormal in the interval $[-1, 1]$, the 16 fitting coefficients describing

the impurity profile can be described uniquely as follows:

$$a_i = \frac{\int_{-1}^1 C(z') L_i(z') dz'}{\int_{-1}^1 L_i^2(z') dz'} \quad (3)$$

$$b_1 = \frac{2}{z_{\min} + z_{\max}} \quad (4)$$

and

$$b_2 = \frac{z_{\max} - z_{\min}}{z_{\min} + z_{\max}} \quad (5)$$

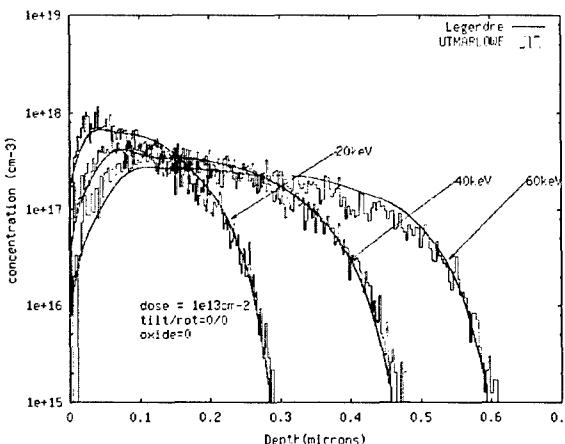


Figure 1. A comparison of the analytical model results with UT-MARLOWE profiles for on-axis low dose Ar implants at three different energies. The interpolation for Ar was carried out between P and As.

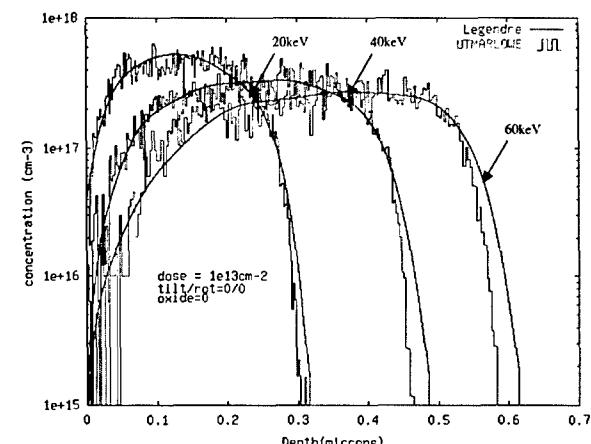


Figure 3. A comparison of the analytical model results with UT-MARLOWE profiles for on-axis low dose F implants at three different energies. The interpolation for F was carried out between B and P.

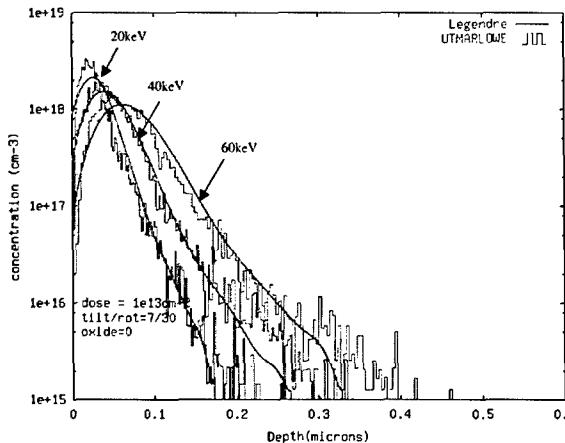


Figure 2. A comparison of the analytical model results with UT-MARLOWE profiles for off-axis low dose Ar implants at three different energies. The interpolation for Ar was carried out between P and As.

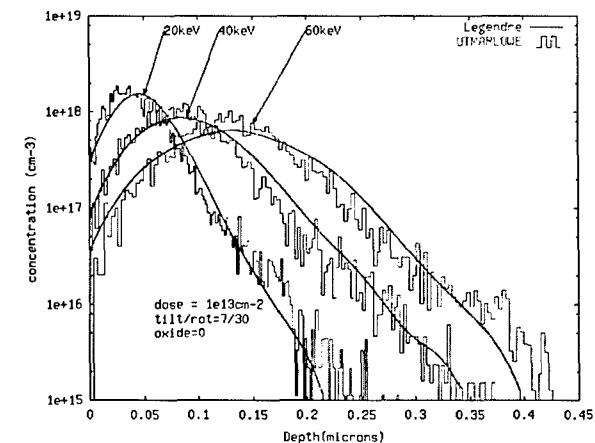


Figure 4. A comparison of the analytical model results with UT-MARLOWE profiles for off-axis low dose F implants at three different energies. The interpolation for F was carried out between B and P.

The fitting coefficients were pre-computed for a small set of implant parameters from the as-implanted impurity profiles generated using UT-MARLOWE using Eqs. (1)–(3). In order to simulate implants for any other set of implant conditions, the coefficients were calculated by interpolating over the various parameters. It was found empirically by trial and error that a square root interpolation of the form, $y = y_1 + (y_2 - y_1) \frac{\sqrt{x} - \sqrt{x_1}}{\sqrt{x_2} - \sqrt{x_1}}$, gave sufficiently accurate results while interpolating over energies, tilt angle, rotation angle and oxide, nitride or silicide layer thickness. While interpolating between two doses, the log of the dose is used

for interpolation and the resulting profile is re-scaled to ensure the correct dose.

In order to extend this approach to model implants of any arbitrary species, two nearest species in terms of atomic mass, one on the left and another on the right of the implanted species, in the periodic table, for which the look-up tables already exist are used. The Legendre coefficients for these two species are computed for the given set of implant parameters using the approach outlined above. The Legendre coefficients for the desired species are then calculated by interpolating between the coefficients of these two species. For implant doses

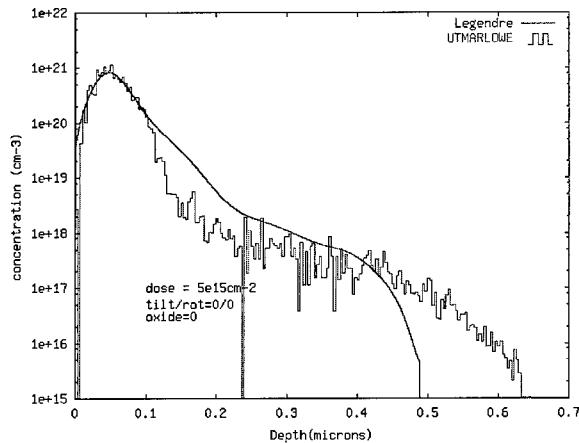


Figure 5. A comparison of the analytical model results with UT-MARLOWE profiles for on-axis high dose Mn implant shows that the simple square root interpolation does not work at high doses. The interpolation for Mn was carried out between P and As.

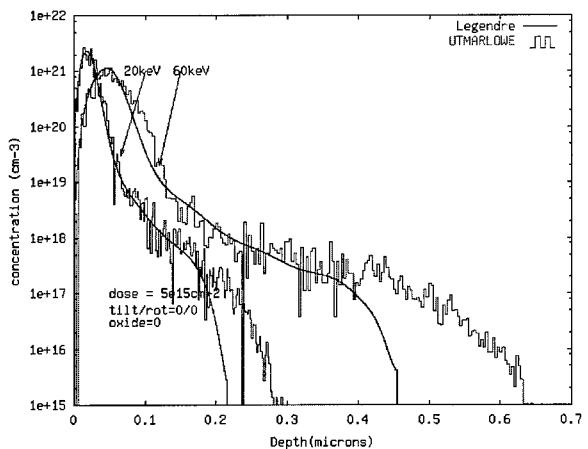


Figure 7. A comparison of the analytical model results with UT-MARLOWE profiles for on-axis high dose Mn implants with the modified interpolation scheme. The results show a good agreement between UT-MARLOWE and the analytical model.

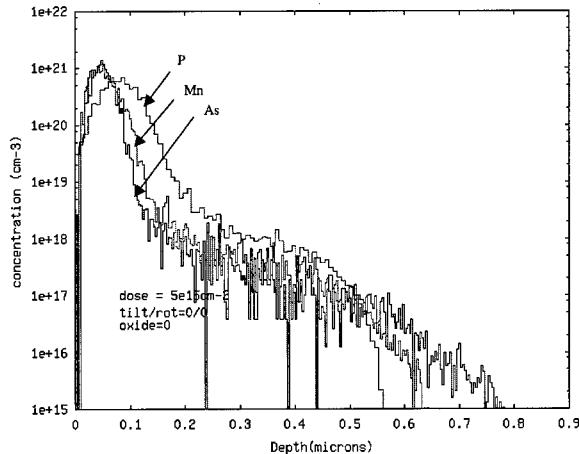


Figure 6. A comparison of UT-MARLOWE profiles for P Mn and As at a dose of $5 \text{ e } 15 \text{ cm}^{-2}$ shows that the Mn profile is much closer to As as compared to P.

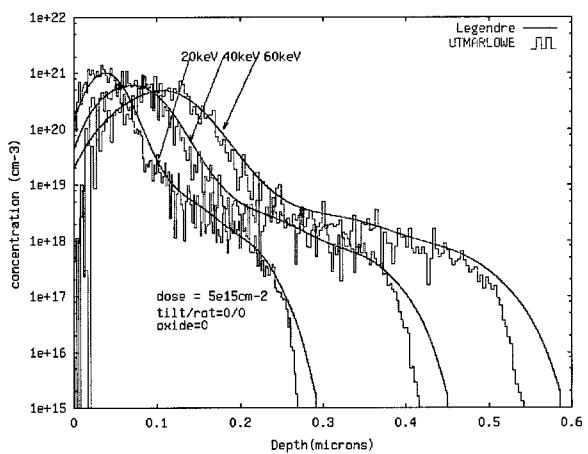


Figure 8. A comparison of UT-MARLOWE profiles with the analytical model predictions for on-axis Mg implants with the modified interpolation scheme. The interpolation for Mg was between B and P.

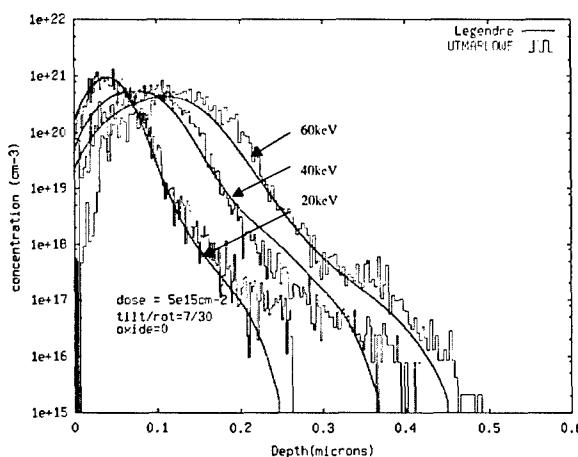


Figure 9. A comparison of UT-MARLOWE profiles with the analytical model predictions for high dose off-axis F implants. The interpolation for F was between B and P.

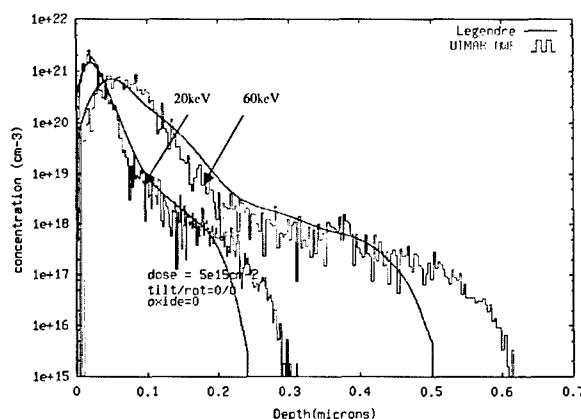


Figure 10. A comparison of UT-MARLOWE profiles with the analytical model predictions for high dose on-axis Ar implants. The interpolation for Ar was between P and As.

less than about $5 \times 10^{13} \text{ cm}^{-2}$, it was found that the square root interpolation gave the best results. To verify the model, it was compared with UT-MARLOWE. Figures 1–4 show some typical results.

For higher doses, as Fig. 5 shows, the square root interpolation is not so accurate. UT-MARLOWE simulations indicate that the damage buildup increases super-linearly with the atomic mass. As a result, the profile of an intermediate species tends to lean more towards the profile of its heavier neighbor. For instance,

Fig. 6 shows a comparison of P, Mn and As profiles, for ion implantation under identical conditions. The Mn profile is extremely close to the As profile.

To account for this, an interpolation function as shown in Eq. (6) was used.

$$y = y_1 + (y_2 - y_1) \frac{f(x) - f(x_1)}{f(x_2) - f(x_1)} \quad (6)$$

where $f(z) = \alpha^z$, z is the atomic number and α is a number between 0 and 1.

If the implant species lies between B and P, α is calculated as

$$\alpha = 1 - 0.1 \log(\text{dose}/5 \times 10^{13}) \text{ cm}^{-2} \quad (7)$$

For a species between P and As, $\alpha = 0.8$ was found to give the best fit over a large range of implant parameters. Figures 7–10 show some typical results.

Conclusion

We have presented an analytical model, which can predict as-implanted profiles for any arbitrary species into crystalline Si. The model has been verified by comparing it extensively with UT-MARLOWE profiles. The comparisons show that the analytical model results are in good agreement with UT-MARLOWE.

This work was supported in part by the Semiconductor Research Corporation.

References

- Balamurugan G., Obradovic B.J., Wang G., Chen Y., and Tasch A.F. 1998. New analytic models and efficient parameter extraction for computationally efficient 1-D and 2-D ion implantation modeling. International Electron Devices Meeting Technical Digest, pp. 517–520.
- Chen Y. 2001. A universal model of ion implantation into topographically complex structures with multiple materials. Ph.D. Thesis, The University of Texas at Austin.
- Morris S.J., Obradovic B.J., Yang S.-H., and Tasch A.F. 1996. Modeling of boron, phosphorus and arsenic implants into single crystal silicon over a wide energy range. International Electron Devices Meeting Technical Digest, pp. 721–724.
- Tasch A.F., Yang S.-H., and Morris S.J. 1995. Modeling of ion implantation into single crystal silicon. Nuclear Instruments & Methods in Physics Research B 102: 173–179.



On the Electron Transient Response in a 50 nm MOSFET by Ensemble Monte Carlo Simulation in Presence of the Smoothed Potential Algorithm

GABRIELE FORMICONE*

Motorola, Inc., 2100 East Elliot Road, Tempe, AZ 85284, USA

gabriele.formicone@motorola.com

MARCO SARANITI

Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL 60616, USA

DAVID K. FERRY

Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University, Tempe, AZ 85287, USA

Abstract. The inclusion of a smoothed potential algorithm within the Ensemble Monte Carlo method (EMC) to account for quantization effects in the inversion layer of a silicon n-MOSFET has been discussed by several authors. Most of the data reported deal with steady state terminal current, transconductance, and capacitance. Within this approach, the electric field acting on each particle is computed from the smoothed potential, which introduces a potential barrier underneath the gate region that pushes the carriers away from the interface, thus accounting for space quantization effects. However, in the EMC method, the electric field at the interface is also used to compute the displacement charge/current during the transient regime. In the implementation of the smoothed potential algorithm, care must be taken when computing this component of the total gate charge. We distinguish between two differently computed electric fields, one from the smoothed potential used for the charge transport and the other one computed from the real potential, as obtained from the solution of Poisson's equation, and used for the displacement charge. We propose this procedure in order to properly include space quantization effects, and at the same time avoid the inaccuracy introduced by the smoothed potential in the displacement charge.

Keywords: Monte Carlo, MOSFET, smoothed potential, transient, simulation

Introduction

State-of-the-art semiconductor technology has already reached dimensions down to the $0.07 \mu\text{m}$ gate length. Several off-the-shelf chips are manufactured using a $0.13 \mu\text{m}$ CMOS process. MOSFETs with gate lengths of order 20 nm have been fabricated in laboratories around the world (Chau *et al.* 2000, Ono *et al.* 1995). Therefore, device simulators must account for, and predict, phenomena that often require the inclusion

and understanding of advanced features. The Ensemble Monte Carlo (EMC) method is believed to be the most accurate approach to simulate carrier transport in the deep sub-micron regime, where non-stationary transport and high electric fields dominate the device properties. Several semiconductor device simulators have been developed that couple the EMC method for the treatment of carrier transport to the self-consistent solution of Poisson's equation (Hockney and Eastwood 1988, Ravaioli and Ferry 1986, Ravaioli *et al.* 2000). In some treatments, the Schrödinger equation has been added to the Poisson's equation for a 2D quantum

*To whom correspondence should be addressed.

mechanical treatment of the electron energy level quantization in the confined silicon inversion layer (Fischetti and Laux 1993, Vasileska and Ferry 1997, Vasileska *et al.* 1995, Jallepalli *et al.* 1997).

Even within this approach, there are some limitations. In fact, the EMC method is a semi-classical approach where scattering mechanisms are treated in a quantum mechanical framework and the transport during the free flights between two scattering events is treated from a classical point of view. In a semiconductor device, the electric field acting on the charged particles is obtained by solving the Poisson equation inside the device. In this framework, the wave-like nature of the carriers is neglected. There have been several studies showing that with channel lengths in the $0.1 \mu\text{m}$ range and below, source-to-drain and gate tunneling, and correlation effects are an important part of the transport process (Ferry 2000a). The confining electrostatic potential in the 2D silicon inversion channel is now competing with the lateral source-to-drain potential barrier where tunneling of the wave-packet associated with the carriers interferes with the drift motion from the source to the drain of a MOSFET. None of these effects are treated by most of the EMC simulators. Ferry, Akis and Vasileska (2000) have shown that the introduction of a smoothed, or effective, potential in the EMC method can account for these quantum-mechanical features. For instance, the smoothed potential accounts for a proper spatial distribution of electrons in the inversion layer by pushing the carriers away from the Si/SiO₂ interface (Vasileska, Schroder and Ferry 1997). In the past, there have been several studies to investigate similar effects in MESFETs by use of the hydrodynamic model (Ancona *et al.* 2000, Zhou and Ferry 1993, 1992a, b, c, Ferry and Zhou 1993).

Previously, we have developed an approach using the smoothed potential algorithm that, without the need to solve Schrödinger equation, can correctly reproduce the main properties of the electron transport in the silicon inversion layer (Formicone *et al.* 2002). However, in the implementation of the smoothed potential algorithm, care must be taken when computing the gate charge. Our goal here is to extend the work done in Formicone *et al.* (2002) and study how the smoothed potential affects the computation of the gate charge in the silicon inversion channel of a MOSFET. Before discussing our results, we will briefly review the EMC model for electron transport in the silicon inversion layer, and the implementation of the smoothed algorithm.

Self-Consistent EMC Model with the Smoothed Potential Algorithm

The Ensemble Monte Carlo model is based on the usual Si band structure for three-dimensional electrons in a set of six nonparabolic Δ valleys (Ferry 2000b). For the silicon inversion layer in the case of a MOSFET, we add surface roughness and impurity scattering for each valley (Formicone *et al.* 2002). Two high-energy phonon scattering are considered via the usual zero-order process, and two low-energy phonon scattering are treated via a first order process (Ferry 2000b).

Impurity scattering is treated within the Born approximation, with a dynamic screening model valid for degenerate and non-degenerate semiconductors. The dynamic screening is introduced through the dielectric function, given by Ferry (2000b):

$$\epsilon(\mathbf{q}, 0) = \epsilon(0) \left[1 + \frac{q_D^2}{q^2} F(\xi, \mu) \right].$$

Here, $q_D = \sqrt{n e^2 / \epsilon_S k_B T}$ is the inverse of the Debye-Hückel screening length and q is the scattered wave vector. The function $F(\xi, \mu)$ is defined as Ferry (2000b):

$$F(\xi, \mu) = \frac{1}{\sqrt{\pi} \xi F_{\frac{1}{2}}(\mu)} \int_0^\infty \ln \left| \frac{x + \xi}{x - \xi} \right| \frac{x dx}{1 + e^{x^2 - \mu}},$$

with $\mu = E_F / k_B T$, $\xi^2 = \hbar^2 q^2 / 8 m k_B T$ and $x^2 = \hbar^2 k^2 / 2 m k_B T = E_k / k_B T$. $F_{\frac{1}{2}}$ is the Fermi integral of order $\frac{1}{2}$. This function can be numerically evaluated and tabulated for different values of the reduced Fermi level μ . In the limit of the Debye screening model, which is valid for lightly doped samples and small q , the screening function F is equal to 1. We use the above formulas for numerically computing the impurity scattering rates, as given by:

$$\Gamma_{imp}(k) = \frac{N_{imp} Z^2 e^4 (2m^*)^{\frac{3}{2}} (1 + 2\alpha E_k) \sqrt{E_k (1 + \alpha E_k)}}{4\pi \hbar^4 \epsilon_S^2} * \int_{-1}^1 \frac{d(\cos \theta)}{\left[2k^2 (1 - \cos \theta) + q_D^2 F(\xi(\cos \theta), \mu) \right]^2}.$$

The term ξ depends on the angle θ since $q^2 = 2k^2(1 - \cos \theta)$ (Formicone *et al.* 2002).

In modeling the Si inversion layer, the 2D nature of the carrier transport is modeled through a hybrid approach where we use our 3D EMC model described

above. To this we add the quasi-2D surface roughness as another scattering process. The surface roughness scattering rate Γ_{SR} is given by Formicone *et al.* (2002) and Ferry (2000b):

$$\Gamma_{SR}(k) = \frac{\pi m^* \Delta^2 L^2 e^2}{2\epsilon_S \hbar^3} e^{\frac{k^3 L^2}{2}} I_0\left(\frac{k^3 L^2}{2}\right) E_{eff}^2$$

Here, I_0 is the modified Bessel function of order 0, k is the magnitude of the in-plane component of the total electron wave vector. E_{eff} is the effective normal electric field computed as $E_{eff} = e(N_{depl} + 0.5n_{inv})/\epsilon_S$, Δ is the average displacement of the interface, and L is the correlation length of the roughness parallel to the interface. We use $\Delta = 0.3$ nm and $L = 2$ nm, in close agreement with experiments. The scattering process is assumed to be purely isotropic in the plane, with angles selected randomly between 0 and 2π . A specular reflection is used for the component of the electron wave vector normal to the surface, thus we are not accounting for DC leakage gate current. Since the normal electric field is position-dependent, in the implementation of the surface roughness scattering we adopt a rejection technique.

The basic steps involved in a self-consistent EMC device simulator consist of the assignment of the charge to the mesh, the solution of Poisson's equation to compute the potential and electric field on the mesh, and the interpolation of the force acting on each particle (Hockney and Eastwood 1988, Jacoboni and Lugli 1989). Our approach adds another step with the computation of a smoothed potential, from which the electric field is derived.

The smoothed potential is obtained from weighting the starting potential with a gaussian function (Ferry 2000a, Ferry, Akis and Vasileska 2000), which in two dimensions reads as:

$$V_{smoothed}(\mathbf{u}) = \int V(\mathbf{u} + \mathbf{t}) G(\mathbf{t}, a_x, a_y) dt$$

G is a gaussian distribution function. From a physical standpoint, it is related to the wave-packet distribution function. The standard deviations in the x and y directions have a different physical meaning (Ferry 2000a). In the source-to-drain (x) direction, the spreading of the wave-packet function is related to the De Broglie wavelength of the electron. For degenerate semiconductors, the De Broglie wavelength is replaced by the Fermi wavelength, thus introducing a substrate doping concentration in the lateral spreading of the electron

wave-packet. In the normal (y) direction, the electron wave distribution depends on the vertical confinement, which depends on the normal electric field. For inhomogeneous situations, like in a semiconductor device, an average value can be used (Ferry 2000a). We use 0.5 nm for the value of a_x , and 1 nm for a_y . \mathbf{t} and \mathbf{u} are two vectors in the x - y plane. At self-consistency, the steady-state current is calculated through the net number of particles crossing one contact per unit time. In fact, the charge flowing through each terminal is recorded at each time step. The charge entering or exiting a contact region consists of both the number of particles that traverse the boundary of the device and the displacement charge arising from the component of the electric field normal to the boundary. Plotted versus time, the slope of the Q-t curve gives, after the initial transient, the current flowing through each terminal. By performing several computer runs, we construct the current-voltage characteristics of the device under test (DUT).

We simulate a MOSFET with a 50 nm gate-length. Details about the simulated structure can be found in Formicone *et al.* (2002).

Results and Discussions

The smoothed potential algorithm computes the electric field from an 'artificial' electric field distribution inside the transistor. This 'artificial' electric field is responsible for the 'correct' spatial carrier distribution in the silicon inversion layer, however it would generate incorrect displacement current at the gate contact. Figures 1 and 2 show our results for the terminal charge in a 50 nm n-MOSFET, computed with and without smoothing of the potential. The data refer to two different bias configurations. The substrate charge is not affected. Minor differences are observed in the source and drain contacts. However, the gate charge shows extremely different results. The same observations can be drawn from analyzing Figs. 3 and 4, where we plot the currents extracted from the previous two figures. In order to explain these differences, we have to look at Fig. 5. The potential and effective potential vertical profiles are shown at mid-gate length. Notice the different slope of the potential at the interface, responsible for the different electric field values computed.

Measuring the current by fitting a straight line, to the data after the first 0.5 ps transient, is a very 'clean' procedure, without noise in the data. In Figs. 3 and 4 the data look very noisy, especially as we approach steady

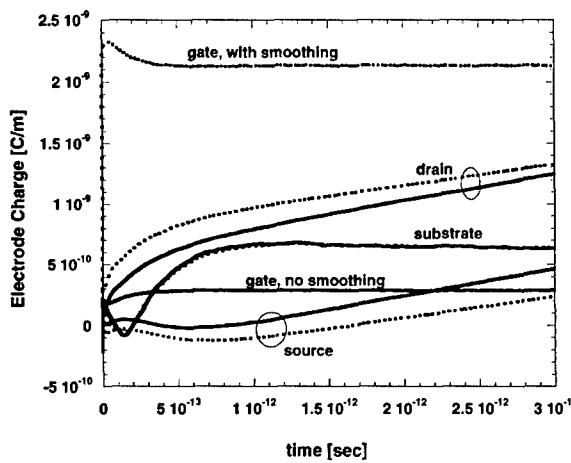


Figure 1. Electrode charge versus time at $V_{ds} = 1$ V and $V_{gs} = 1$ V with and without the smoothing of the potential. Solid lines are for data without smoothing; dashed lines include the smoothing of the potential.

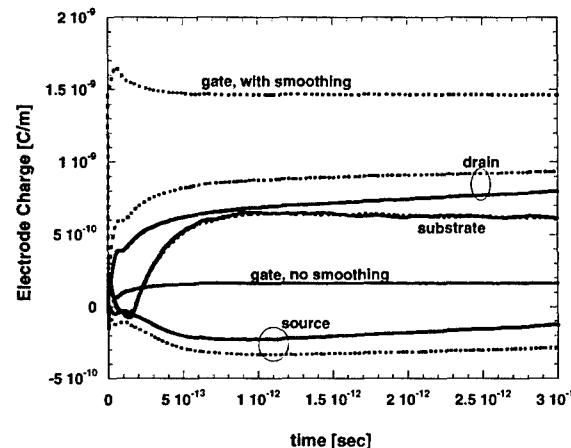


Figure 2. Electrode charge versus time at $V_{ds} = 1$ V and $V_{gs} = 0.5$ V with and without the smoothing of the potential. Solid lines are for data without smoothing; dashed lines include the smoothing of the potential.

state. These two curves have been obtained by smoothing the original data set with a 50-point average. Notice that, when the transistor is operated near threshold, it actually undergoes current excursions twice as large as when the device is in transconductance saturation. Also, the transistor is showing two different transient regimes, regardless of its bias conditions. One regime lasts less than 0.1 ps. This is the time during which the drain contact region is building up a depletion region in response to the external voltage excitation. The second transient regime is observed from 0.1 ps to about 0.5 ps.

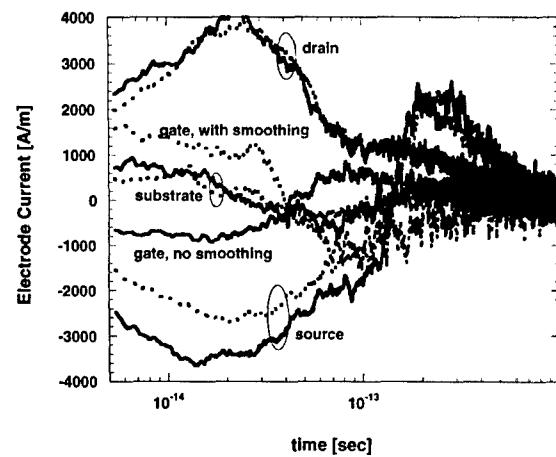


Figure 3. Electrode current versus time at $V_{ds} = 1$ V and $V_{gs} = 1$ V with and without the smoothing of the potential. Solid lines are for data without smoothing; dashed lines include the smoothing of the potential.

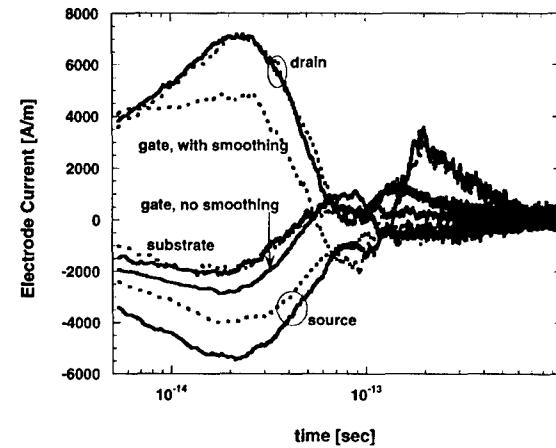


Figure 4. Electrode current versus time at $V_{ds} = 1$ V and $V_{gs} = 0.5$ V with and without the smoothing of the potential. Solid lines are for data without smoothing; dashed lines include the smoothing of the potential.

This is the time it takes to build up enough charge in the channel to achieve inversion. This is the device transient response that reflects the carrier transient dynamic behavior in response to the externally applied voltages.

Based on the above observations, we make a distinction between the particle and displacement electric fields, and use them appropriately in order to get correct results. As a matter of fact, it would be extremely inconsistent to use the terminal charges if charge conservation is not guaranteed. EMC is a device simulation technique especially suitable for transient simulations.

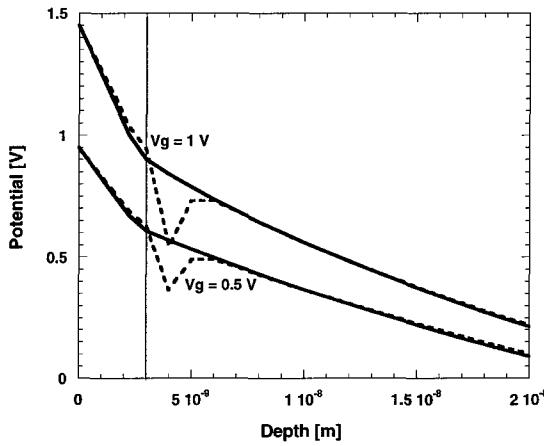


Figure 5. Vertical profile of the electrostatic potential with and without smoothing, for $V_{ds} = 1$ V, at $V_{gs} = 1$ V and $V_{gs} = 0.5$ V. The profile is taken at half the gate length. Notice the 3 nm gate oxide at the left of the vertical line in the plot.

At the same time, the transient response of a transistor is strongly dependent on the intrinsic device capacitances. If used with lumped elements (Contrata *et al.* 2001), for mixed-mode type of simulations, charge conservation is a very important property that must be ensured, and this is accomplished by computing displacement terminal charges from the intrinsic, i.e. not smoothed, potential.

In summary, our analysis has shown that all the advantages that the smoothed potential algorithm brings to semiconductor device simulation can be retained when properly computing the displacement charge.

References

- Ancona M.G. *et al.* 2000. Density-gradient analysis of MOS tunneling. IEEE TED 47: 2310.
- Chau R. *et al.* 2000. 30 nm physical gate length CMOS transistors with 1.0 ps n-MOS and 1.7 ps p-MOS gate delays. IEDM. 45.
- Contrata W. *et al.* 2001. Intermodulation distortion simulation using physical GaAs FET model. International Microwave Symposium.
- Ferry D.K. 2000a. Effective potentials and the onset of quantization in ultrasmall MOSFETs. Superlattices and Microstructures 28: 419.
- Ferry D.K. 2000b. Semiconductor Transport. Taylor & Francis.
- Ferry D.K., Akis R., and Vasileska D. 2000. Quantum effects in MOSFETs: Use of an effective potential in 3D Monte Carlo simulation of ultra-short channel devices. IEDM Tech. Dig. IEEE Press, New York, p. 287.
- Ferry D.K. and Zhou J.-R. 1993. Form of the quantum potential for use in hydrodynamic equations for semiconductor device modeling. Physical Review B 48: 7944.
- Fischetti M.V. and Laux S.E. 1993. Monte Carlo study of electron transport in silicon inversion layers. Physical Review B 48: 2244.
- Formicone G. *et al.* 2002. Study of a 50 nm n-MOSFET by ensemble Monte Carlo simulation including a new approach to surface roughness and impurity scattering in the Si inversion layer. IEEE TED 49: 125.
- Hockney R.W. and Eastwood J.W. 1988. Computer Simulation Using Particles. Institute of Physics Publishing.
- Jacoboni C. and Lugli P. 1989. The Monte Carlo Method for Semiconductor Device Simulation. Springer.
- Jallepalli S. *et al.* 1997. Electron and hole quantization and their impact on deep submicron silicon p- and n-MOSFET characteristics. IEEE TED 44: 297.
- Ono M., Saito M., Yoshitomi T., Fiegnan C., Ohguro T., and Iwai H. 1995. A 40 nm gate length n-MOSFET. IEEE TED 42: 1822.
- Ravaioli U. and Ferry D.K. 1986. MODFET ensemble Monte Carlo model including the quasi-two-dimensional electron gas. IEEE TED 33: 677.
- Ravaioli U. *et al.* 2000. Monte Carlo simulation of ultra-small MOS devices. Superlattices and Microstructures 27: 137.
- Vasileska D., Bordone P., Eldridge T., and Ferry D.K. 1995. Calculation of the average interface field in inversion layers using zero-temperature Green's function formalism. J. Vac. Sci. Technol. B 13: 1841.
- Vasileska D. and Ferry D.K. 1997. Scaled silicon MOSFET's: Universal mobility behavior. IEEE TED 44: 577.
- Vasileska D., Schroder D.K., and Ferry D.K. 1997. Scaled silicon MOSFET's: Degradation of the total gate capacitance. IEEE TED 44: 584.
- Zhou J.-R. and Ferry D.K. 1992a. Simulation of ultra-small GaAs MESFET's using quantum moment equations-II: Velocity overshoot, IEEE TED 39: 1793.
- Zhou J.-R. and Ferry D.K. 1992b. Simulation of ultra-small GaAs MESFET using quantum moment equations. IEEE TED 39: 473.
- Zhou J.-R. and Ferry D.K. 1992c. Ballistic phenomena in GaAs MESFETs: Modeling with quantum moment equations. Semicond. Sci. Technol. 7: B546.
- Zhou J.-R. and Ferry D.K. 1993. Modeling of quantum effects in ultrasmall HEMT devices. IEEE TED 40: 421.



Quantum Corrections in the Monte Carlo Simulations of Scaled PHEMTs with Multiple Delta Doping

K. KALNA AND A. ASENOV

*Device Modelling Group, Department of Electronics & Electrical Engineering, University of Glasgow,
Glasgow, G12 8LT, Scotland, UK*

Abstract. The effective potential approach which can represent quantum mechanical (QM) confinement at a heterointerface has been incorporated into our Monte Carlo device simulator MC/H2F. The simulator is used to investigate the impact of the quantum corrections on the performance of single and double δ -doped pseudomorphic high electron mobility transistors scaled to decanano dimensions. The QM confinement in the device channel results in reduction of the drive current and the device transconductance. Its influence increases with the device scaling from 120 to 30 nm gate length and also with increasing the carrier sheet density in the double δ -doped structures.

Keywords: MC device simulations, effective potential, PHEMT, scaling

1. Introduction

The performance of pseudomorphic high electron mobility transistors (PHEMTs) improves as these devices are scaled to deep decanano dimensions (Kalna *et al.* 2000). The scaling has to be done in both lateral and vertical directions of the PHEMTs to achieve an increase in the device transconductance and in RF figures of merit (Kalna *et al.* 2000). However, the carrier density in the device channel decreases with the reduction of the gate-to-channel separation. This decrease can be compensated with the introduction of a second delta doping layer into the vertical device structure (Hur *et al.* 1996). The second delta doping can be placed either below or above the channel, near the gate. The doping level and location of the second delta doping have to be carefully chosen to provide the desired improvements in device performance.

The reduction of the gate-to-channel separation and rapid variations of the electric field in the scaled devices also enhances the impact of quantum mechanical (QM) confinement effects on the device performance. Therefore, we would like to investigate how the QM confinement effects affect the PHEMT performance (Kalna *et al.* 2000) when the device is scaled in respect to gate lengths of 120, 90, 70, and 50 and 30 nm.

Particularly, we compare the influence of the QM confinement effects on the performance of single and double doped structures. In the second case we consider possible placements of the second delta doped layer above or below the channel. We have carried out the simulations in which we have implemented quantum corrections using the effective potential (EP) formalism (Ferry 2000a).

The presence of an additional delta doping layer can increase the carrier sheet density in the channel up to $6 \times 10^{12} \text{ cm}^{-2}$. Several authors claim that at such high carrier densities the effect of degeneracy (Fermi exclusion principle) has to be included in scattering events which can account, to some extent, for the transport enhancement in quantum wells (Mateos *et al.* 2000). Therefore, we have also included the degeneracy in the MC module of the device simulator in order to faithfully investigate its possible influence on electron transport in the scaled PHEMTs.

2. Effective Potential in Device Simulations

The MC/H2F simulator uses a finite element approach to couple the Poisson equation to particle motion in a complicated PHEMT geometry, which involves the

T-shape gate and the recess regions. The MC module of MC/H2F includes electron interactions with polar optical phonons, inter- and intra-valley non-polar optical phonons, and acoustic phonons. Scattering with ionized and neutral impurities is also considered. Moreover, the alloy scattering and the effect of strain (Köpf, Kosina and Selberherr 1997) on the bandgaps, the effective masses, and the polar optical phonon energies and potentials are taken into account in the InGaAs channel. All scattering rates, and their final states, are calculated within a form factor (overlap integral) given by $F(E, E') = [(1 + \alpha E)(1 + \alpha' E') + 1/3\alpha E \alpha' E'] / (1 + 2\alpha E)(1 + 2\alpha' E')$ (Kane 1957). In this formula, an electron with an initial energy E scatters into the state with a final energy E' . α and α' are non-parabolicity parameters for the electron in initial and final valleys, respectively. The implementation of the above form factor into the actual analytical non-parabolic band structure allows us to extend the validity of the MC/H2F electron transport model up to 400 kV/cm. Further, the MC device simulator has been calibrated against a real 120 nm gate length PHEMT fabricated at our department. Simulated data which have been already remapped to include external resistances are in excellent agreement with experimental measurements (Kalna *et al.* 2000).

The QM confinement effects which appear at the AlGaAs/InGaAs heterointerface of PHEMTs may be introduced at low cost in MC simulations using the EP approach (Ferry *et al.* 2000a). The EP smooths the classical potential and accounts for quantum confinement effects like displacement of the mobile charge centroid away from the heterojunction interface between the channel and the supply layer. The EP approach has been adapted for the MC/H2F as follows. During each time step, required by self-consistent MC simulations (Hockney and Eastwood 1988), a classical potential P_{class} , obtained by solving the Poisson equation, is smoothed by Gaussian distribution $G(\mathbf{x})$ to get an EP, P_{eff} , as (Ferry *et al.* 2000a)

$$P_{\text{eff}} = \int d\mathbf{r}' P_{\text{class}}(\mathbf{r} + \mathbf{r}') G(\mathbf{r}'), \quad (1)$$

$$G(\mathbf{x}) = \frac{1}{a\sqrt{2\pi}} \exp\left(-\frac{x^2}{2a^2}\right), \quad (2)$$

where a is the standard deviation that may be approximated by $a = \hbar/\sqrt{12m k_B T}$ (Ferry 2000b). However, the standard deviation of the Gaussian (2) is adjusted to match a difference between the classical Poisson

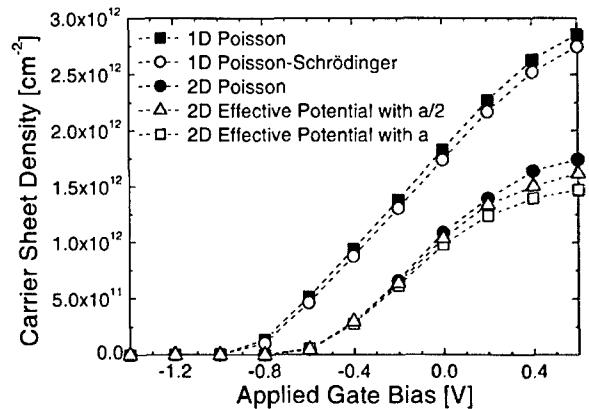


Figure 1. Carrier sheet density as a function of the gate bias for the 120 nm single doped PHEMT which is used to calibrate the standard deviation in Eqs. (1) and (2). The shift between the 1D classical solution obtained from Poisson's equation only and the 1D self-consistent solution from coupled Poisson-Schrödinger equation is reflected in the 2D models.

solution and the self-consistent Poisson-Schrödinger solution considering a 1D model of the PHEMT structure. Figure 1 illustrates this procedure for the 120 nm gate length single doped PHEMT. A variation between the carrier sheet density obtained using the 2D EP with a standard deviation $a_{\text{cal}} = a/2$ and the carrier sheet density obtained using the 2D classical potential is much closer to a variation between the 1D classical and 1D self-consistent solutions. Therefore, the standard deviation a_{cal} has been applied in Eq. (2) for the PHEMTs. P_{eff} is then used in the propagation routine of the MC module at each time step. A smoothing of P_{class} by the Gaussian (2) is performed only for the electron transport in the Γ valley. We assume that the electron transport in higher conduction valleys, L and X , is not affected by a carrier confinement at the AlGaAs/InGaAs heterointerface (Park and Brennan 1990) because these valley are energetically much higher than the Γ valley band offset. Moreover, most of the carriers which travel through the InGaAs channel are in the Γ valley anyway.

3. Impact on PHEMT Performance

The investigated PHEMT has a T-shaped gate and an InGaAs channel with the indium content of 0.2. The active doping concentration of the original delta doped layer is assumed to be $3.5 \times 10^{12} \text{ cm}^{-2}$. The conduction band profile of the 120 nm double doped PHEMT when a second delta doping layer is placed below the

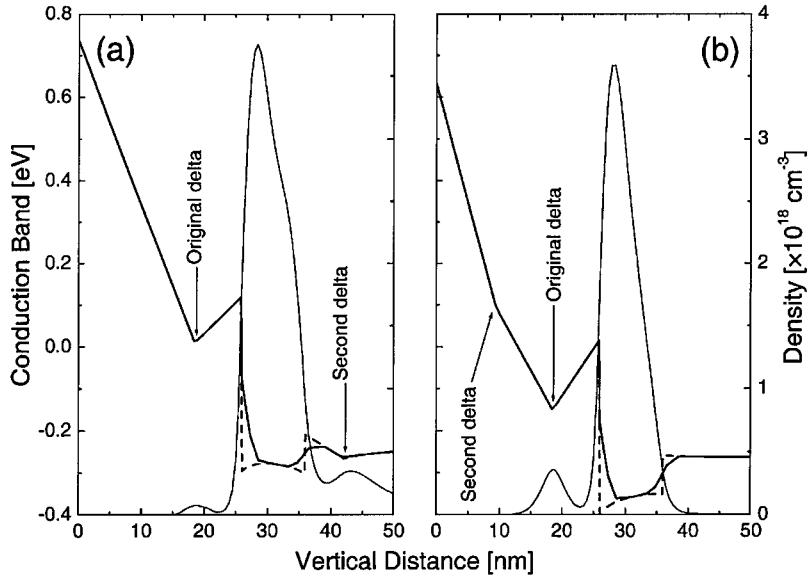


Figure 2. Conduction band profile and carrier density as functions of the depth in the 120 nm double doped PHEMT with the second delta doped layer (a) below the channel or (b) above the original doping. The classical profile is shown by the dashed line and the smoothed EP profile by the full line.

channel can be seen on Fig. 2(a) together with the profile smoothed by using the EP. Figure 2(b) shows the conduction band profile if an additional delta doping layer is placed above the original doping, near to the gate, in order to screen high electric fields (due to the gate fringing effect) and to maintain a better control of the device channel. The smoothed EP profile is shown as well. The increase in carrier densities for both additional delta doping placements is also apparent from Fig. 2. The placement of the second delta doping below the channel results in substantial increase of the device current and in broadening of the transconductance peak which will result in a dramatic improvement in the device linearity. Placing the second delta doping above the channel does not improve the device linearity too much but can remarkably increase the device transconductance.

We have monitored an average carrier velocity along the InGaAs channel when a MC simulation includes or excludes the EP. Figure 3 shows that the average carrier velocity in both single doped 120 and 70 nm gate length PHEMTs is slightly larger when the EP corrections are applied. The effect of the quantum confinement is more pronounced around the peaks of the velocities.

The quantum confinement has a detrimental effect on both single and double doped PHEMTs because it shifts the channel charge centroid away from the gate.

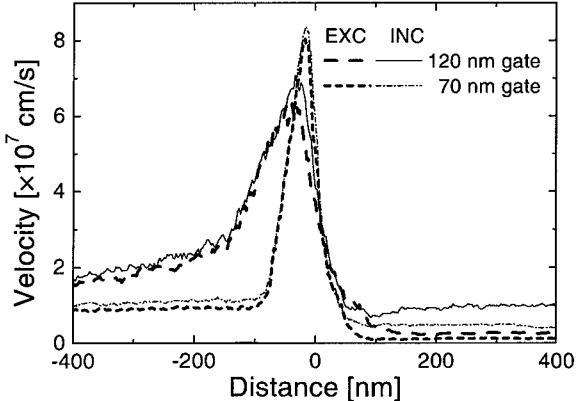


Figure 3. Average carrier velocity along the InGaAs channel for the 120 and 70 nm single doped PHEMTs at gate and drain voltages of 0.0 V and 1.5 V, respectively. The EP is excluded (EXC) or included (INC) in simulations.

The result is that the drive current and the transconductance of the devices decrease as shown in Figs. 4–6. As the devices are scaled from 120 nm gate length to 70 and 50 nm gate lengths the impact of the QM confinement increases (Fig. 4) because of a large relative increase in the gate to channel separation. The placement of the additional delta doped layer above the original doping in the PHEMT structure can increase the device

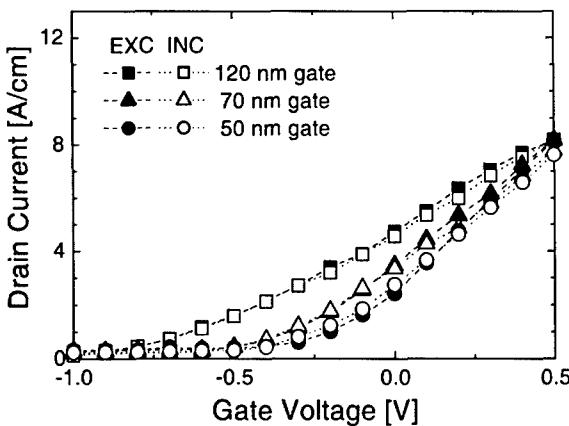


Figure 4. I_D - V_G characteristics (symbols) and transconductances (lines) for intrinsic devices with the single delta doping layer at a drain bias of 1.5 V. Quantum corrections using the EP approach are excluded (open symbols) or included (full symbols) in simulations.

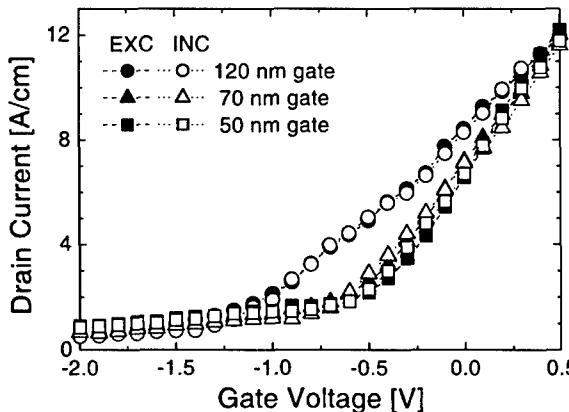


Figure 5. I_D - V_G characteristics (symbols) and transconductances (lines) for intrinsic double delta doped devices at a drain bias of 1.5 V. The second delta doping layer is placed below the channel. Again, quantum corrections are excluded (open symbols) or included (full symbols) in simulations.

transconductance by approximately 50% in the 120 nm PHEMT and up to nearly 80% in the 70 nm PHEMT as shown in Figs. 5 and 6. Below 50 nm gate lengths the increase in the intrinsic transconductance is less pronounced due to the small relative distance between the additional and the original delta doping layers. Figure 6 also shows that the EP reduction in the drive current is more pronounced in the double doped structures compared to the single doped ones. This is particularly evident for the 30 nm double doped PHEMT with the second delta doping near the gate which has its transconductance reduced by 15%.

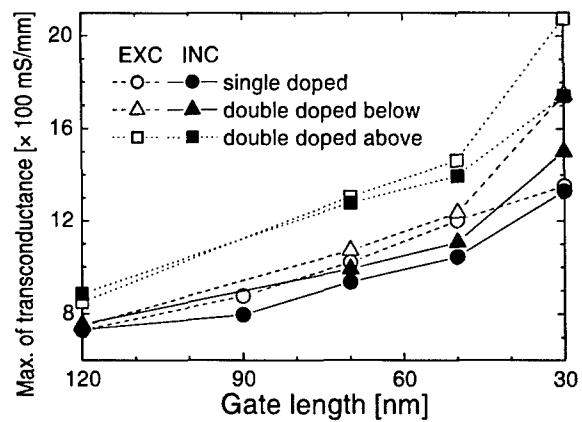


Figure 6. Maximum of transconductance versus the gate length of investigated PHEMTs. A smoothing by the EP is excluded (EXC) or included (INC).

4. Conclusions

Extensive MC device simulations of single and double doped PHEMTs scaled into decanano dimensions have been carried out to study a possible improvement in the device linearity and/or transconductance. The QM confinement effects are included in simulations using the EP approach (Ferry 2000a). The standard deviation of Gaussian used to smooth the classical potential in EP simulations was calibrated against the self-consistent Poisson-Schrödinger solution. When the EP is used to calculate electric fields in the devices the drive current and the transconductance degrade reflecting the fact that the QM confinement shifts the channel charge centroid away from the gate. This causes the loss of gate control over the device channel. The QM confinement effects become more pronounced with the device scaling from the 120 to 30 nm gate length and also with increasing of the device sheet density in the double doped structures.

We have also incorporated degeneracy into the MC module of MC/H2F using an approach suggested in (Fischetti and Laux 1988). Nevertheless, we do not observe any noticeable influence of degeneracy (Mateos *et al.* 2000) on the scaled PHEMTs whether single or double doped.

References

- Ferry D.K. 2000a. *Superlatt. Microstruct.* 27: 61.
- Ferry D.K. 2000b. In: Proceeding of International Workshop on Computational Electronics (IWCE-7) Glasgow, Barker J. (Ed.), p. 63.
- Fischetti M.V. and Laux S.E. 1988. *Phys. Rev. B* 38: 9721.

- Hockney R.W. and Eastwood J.W. 1988. Computer Simulation Using Particles. Adam Hilger, Bristol.
- Hur K.Y., Hetzler K.T., McTaggart R.A., Vye D.W., Lemonias P.J., and Hoke W.E. 1996. Electron. Lett. 32: 1516.
- Kalna K., Roy S., Asenov A., Elgaid K., and Thayne I. 2000. In: Lane W.A., Crean G.M., McCabe F.A., and Grünbacher H. (Eds.), Proceedings of ESSDERC 2000. Frontier Group, Cork, p. 156.
- Kane E.O. 1957. J. Phys. Chem. Solids 1: 249.
- Köpf C., Kosina H., and Selberherr S. 1997. Solid-State Electron. 41: 1139.
- Mateos J., Gonzales T., Pardo D., Hoel V., Happy H., and Cappy A. 2000. IEEE Trans. Electron Devices 47: 250.
- Park D.K. and Brennan K.F. 1990. IEEE Trans. Electron Devices 37: 618.



Thermally Self-Consistent Monte Carlo Device Simulations

N.J. PILGRIM, W. BATTY AND R.W. KELSALL

*Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering,
University of Leeds, Leeds, UK*

Abstract. We present details of a Monte Carlo simulation code which is coupled to a Heat Diffusion Equation (HDE) solver. Through an iterative procedure, which bypasses the differences in electronic and thermal timescales, this coupled code is capable of producing steady-state thermally self-consistent device characteristics. Electronically-generated thermal flux is calculated by monitoring the net rate of phonon emission, which may be resolved both spatially and by phonon type. The thermal solution is extracted through use of a novel analytical thermal resistance matrix technique which avoids calculation of temperatures beyond the electronically important device region while including the large-scale boundary conditions. On application to a GaAs MESFET the expected ‘thermal droop’ behaviour is obtained in the I-V characteristics and we find a linear relationship between peak lattice temperature and applied source-drain bias. At moderate biases the contribution of intervalley phonons to the thermal power output surpasses that of optical phonons.

Keywords: electrothermal, Monte Carlo, MESFET, III-V, hot phonons

1. Introduction

The Monte Carlo technique has been extensively applied to the modelling of electronic transport in semiconductor devices, yet with the exception of a few results for silicon devices (Yoder and Fichtner 1998, Tarnay *et al.* 1997), Monte Carlo device models have been applied universally under an isothermal approximation. Other electronic models have previously been extended to include thermal self-consistency (Yoder and Fichtner 1998, Johnson, Snowden and Pollard 1997, Houngh *et al.* 2000) and have shown this to be important in reproducing I-V characteristics of group III-V FETs (Atherton, Snowden and Richardson 1993). Group III-V GaAs-based devices are known to be generally more thermally active than their silicon-based counterparts, due to among other factors a $3\times$ lower thermal conductivity and a strong polar channel for phonon emission absent in group IV materials.

In this paper we present details of a self-consistent electrothermal Monte Carlo simulator including, as far as we are aware, the first such modelling of a III-V FET. In the next section we first discuss computational details involved with the construction of the simulator

code. We then examine results from the application of the code to a GaAs MESFET, including thermally self-consistent I-V characteristics.

2. Computational Details

The electronic Monte Carlo model forming the core of this prototype simulator includes standard 3-valley nonparabolic models of the electronic bandstructure of the III-V materials which comprise the device. Electronic scattering associated with ionised impurities and all appropriate phonon interactions are included, with long-range Coulombic interactions accounted for by self-consistent solution of the two-dimensional Poisson equation.

2.1. Scale Issues

Monte Carlo simulations involve a direct temporal evolution of a representative ensemble of particles, typically for total simulation times of the order of tens of picoseconds over an active device cross-sectional area of the order of square microns. On the contrary, thermal diffusion occurs on much larger space- and time-scales,

of the order of nanoseconds to microseconds over thousands of square microns. To obtain a combined electrothermal transient solution would therefore require simultaneous simulation of the electronic and thermal components over the longer thermal timescale. While this presents no problem in obtaining a thermal solution, a corresponding electronic solution using the Monte Carlo method would be computationally infeasible due to the long simulation times which would be required. For this reason we solve only for steady-state solutions, using an iterative method where the electronic and thermal components are solved alternately until convergence is achieved. Convergence is determined by analysis of source-drain currents, but typically $\sim 10\text{--}20$ electrothermal iterations are performed to ensure correctness.

The coupling between the two solvers occurs via the spatially-resolved mean rate of emitted thermal flux generated by the Monte Carlo algorithm and the temperature distribution subsequently generated by the thermal solver (and fed back into the next Monte Carlo evolution). While the minimum area over which the distribution of emitted thermal flux and temperature is required is simply the active region of the device modelled by the Monte Carlo solver, the thermal boundary conditions of the full device die must be included.

2.2. Thermal Solver

A typical choice of thermal solver might involve spatial discretisation of the Heat Diffusion Equation (HDE) using the finite difference or finite element methods. The problem with applying these methods here arises in the need to discretise, and solve for temperatures over, the entire thermal domain when the area of interest is far smaller. In order to account for the large-scale thermal boundary conditions but avoid solving for the temperature distribution anywhere but over the minimal active device region a novel analytical thermal resistance matrix technique (Batty *et al.* 2001) has been used. Thermal nonlinearity due to the temperature dependence of the thermal conductivity is included via appropriate application of the Kirchhoff transformation (Bonani and Ghione 1995). The thermal domain is currently assumed to be a simple cuboid with adiabatic top and side surfaces and a fixed-temperature heat-sink at the base. However this is not a intrinsic limitation of the technique and more complex domains can be constructed to more accurately match recessed device structures.

2.3. Electronic Model Modifications

In more macroscopic models of electronic transport the distribution of emitted thermal flux may be extracted via calculation of the dot-product of the electric field and current density or even more simply using $I_{ds}V_{ds}$. The Monte Carlo solver may also calculate the thermal flux using this method, but in addition is able to obtain a more dynamic estimate by monitoring the net emission rate of phonons over the duration of the simulation.

Since the temperature distribution in all but the first Monte Carlo simulation is non-uniform, electron-phonon scattering rates for a variety of temperature points are required. Various schemes may be applied to deal with this, including complete tabulation of all possible rates at the start of each simulation or tabulation over a range of temperatures with suitable interpolation. Here we have chosen to use the rejection technique as discussed in previous Monte Carlo simulations of nonequilibrium phonons (Rieger *et al.* 1989), where the pre-tabulated rates are calculated with a maximal expected phonon occupation N_{\max} . When an electron is selected to scatter via a phonon process whose rate is calculated using this maximal value, the probability that the process will be accepted as a real scattering event in the simulation is given by

$$P_{real} = \frac{N_{real} + \beta}{N_{\max} + \beta} \quad (1)$$

where β is unity for emission processes and zero for absorption processes. While we include the variation of all electron-phonon scattering processes with temperature, we currently neglect the net rate of intravalley acoustic phonon emission since we are concerned with 300 K operation where this contribution is very small.

3. Results

The simulation code described above has been applied to a 0.2 μm gate GaAs MESFET, considered contained within a $500 \times 500 \times 125 \mu\text{m}$ die. The structure is shown in Fig. 1.

Figure 2 shows the spatial distribution of (isothermal) net phonon emission at source-drain biases of 1 V and 5 V. A sharp peak in the emission occurs near the corresponding high peak in the electric field profile at the edge of the drain contact. These peaks occur downstream (from the perspective of the electrons) of the peak in the electric field, as also seen by Moglestue,

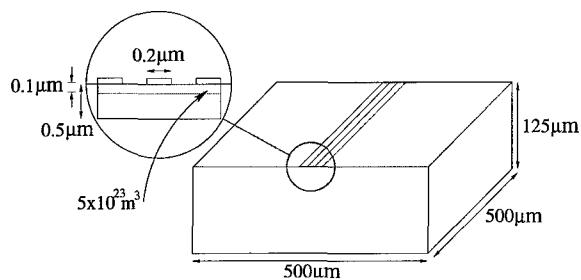


Figure 1. Structure of simulated GaAs MESFET.

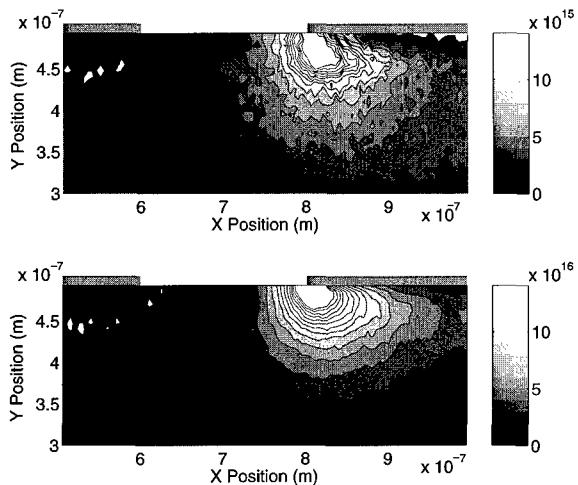


Figure 2. Spatial distribution of net phonon emission at V_{ds} of 1 V (above) and 5 V (below).

Buot and Anderson (1995), occurring just within the gate end of the drain. A low-emission ridge is also visible below the depletion region at 1 V; this is also present at 5 V but is not resolved by the contours.

The change with bias of the relative contribution to the total (isothermal) thermal flux by different phonon types is shown in Fig. 3; we are not aware of an analysis of heat generation in this way in previous work. Categorisation of phonon types is according to their reciprocal-space location: Γ phonons (small- q intravalley processes), L phonons ($\Gamma-L$ and $L-X$ intervalley transitions) and X phonons ($\Gamma-X$, $L-L$ and $X-X$ intervalley transitions). By moderate source-drain biases just beyond approximately 2 V the emission by (optical intravalley) Γ phonons is matched by the sum of the (intervalley) L and X phonon emission. Furthermore beyond just after 4 V the Γ phonon emission is overtaken by the X phonon emission.

On inclusion of thermal self-consistency, source-drain currents show the characteristic ‘thermal droop’

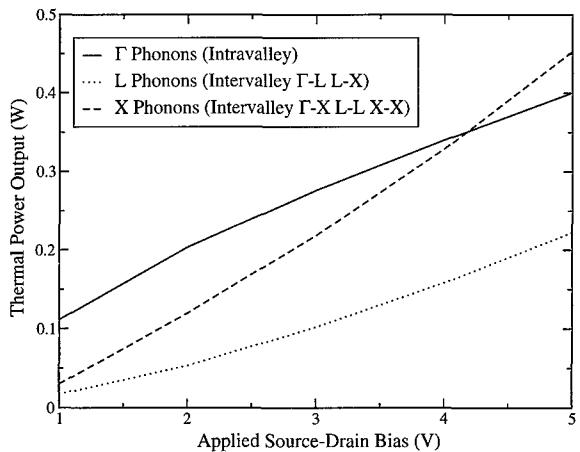


Figure 3. Change with bias of total net phonon emission associated with Γ , L and X phonons.

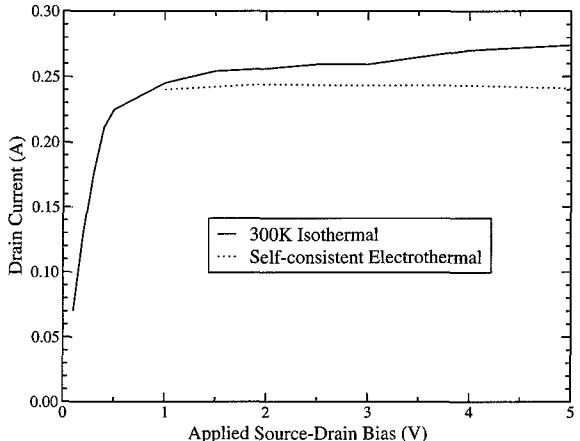


Figure 4. MESFET I-V characteristics, with and without thermal self-consistency.

behaviour. Figure 4 shows the extent of this shift at 0 V (applied) gate bias: at high source-drain biases the slowly increasing isothermal current becomes almost flat, with a slight decrease towards 5 V.

In contrast with the almost constant drain current in the saturation region, the peak temperature present increases in a strongly linear fashion at a rate of $\approx 13.5 \text{ K}/V_{ds}$ over the same range (Fig. 5). Yoder and Fichtner (1998) obtain a similar quasi-linear increase in Si MOSFETs for lower biases, which becomes superlinear at higher biases ($\sim 4 \text{ V}$). However their more strongly peaked spatial temperature distribution suggests a far smaller thermal (die) size was used, which might explain this difference.

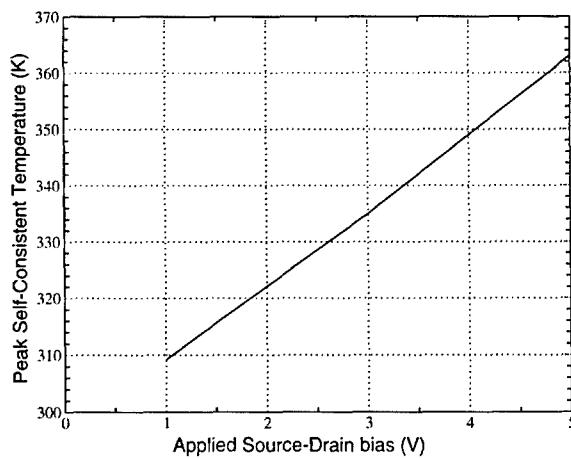


Figure 5. Variation of peak temperature within the MESFET with applied source-drain bias.

4. Conclusion

We have presented details of a Monte Carlo simulation code which is capable of determining steady-state thermally self-consistent device characteristics. Results from the application of this code to a $0.2\text{ }\mu\text{m}$ gate GaAs MESFET with a $500 \times 500 \times 125\text{ }\mu\text{m}$ die are shown. The electronically-generated thermal flux is calculated by counting the net rate of phonon emission. We examine its spatial distribution and the contribution of different phonon types at different source-drain

biases. Thermally self-consistent simulation gives the expected ‘thermal droop’ effect on the I-V behaviour and a linear rise in the peak lattice temperature with source-drain bias.

Acknowledgment

This work is funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC).

References

- Atherton J.S., Snowden C.M., and Richardson J.R. 1993. IEEE MTT-S Digest, pp. 1181–1184.
- Batty W., Christoffersen C.E., David S., Panks A.J., Johnson R.G., Snowden C.M., and Steer M.B. 2001. In: Proceedings 17th Annual IEEE SemiTherm Symp, San Jose, pp. 71–84.
- Bonani F. and Ghione G. 1995. Solid-State Electronics 38(7): 1409–1412.
- Houng M.-P., Wang Y.-H., Chong K.-K., Chu C.-H., Hung C.-I., and Miaw J.-W. 2000. J. Appl. Phys. 88(5): 2553–2559.
- Johnson R.G., Snowden C.M., and Pollard R.D. 1997. In: IEEE MTT-S Internat. Microwave Symp. Dig., Vol. 3, pp. 1485–1488.
- Moglestuc C., Buot F.A., and Anderson W.T. 1995. J. Appl. Phys. 78(4): 2343–2348.
- Rieger M., Kocevar P., Lugli P., Bordone P., Reggiani L., and Goodnick S.M. 1989. Phys. Rev. B 39(11): 7866–7875.
- Tarnay K., Gali A., Poppe A., Kocsis T., and Masszi F. 1997. Physica Scripta. T69: 290–294.
- Yoder P.D. and Fichtner W. 1998. In: de Meyer K. and Biesemans S. (Eds.), *Simulation of Semiconductor Devices and Processes*, Springer, Wien, pp. 165–168.



3D Monte Carlo Modeling of Thin SOI MOSFETs Including the Effective Potential and Random Dopant Distribution*

S.M. RAMEY[†] AND D.K. FERRY

Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University,
Tempe, AZ 85287-5706, USA

steve.ramey@asu.edu

Abstract. We use the effective potential to include quantum mechanical effects in thin SOI MOSFETs simulated with 3D Monte Carlo. We explore the role of discrete dopant distributions on the threshold voltage of the device within the framework of the effective potential by examining the current-voltage behavior as well as the electron distributions within the device. We find that simulations with the effective potential produce a similar shift in current as classical simulations when the dopants are considered to have a random discrete distribution instead of a uniform distribution.

Keywords: effective potential, SOI MOSFET, Monte Carlo

1. Introduction

As modern devices continue to scale to smaller sizes, it has become imperative to include quantum mechanical effects when modeling device behavior. Such effects can in theory be treated by a self-consistent solution of the Schrödinger equation, but this approach has proved difficult to implement in an ensemble Monte Carlo simulation. We have recently proposed the use of the effective potential to treat the quantum mechanical effects of charge set-back from the oxide interface, and the increased ground-state energy of electrons in the inversion layer (Ferry 2000). Furthermore, the importance of including discrete random dopant distributions has been shown for many types of devices (Zhou and Ferry 1995, Gross, Vasileska and Ferry 2000). In this work, we extend the effective potential method to model ultrasmall, SOI MOSFETs with random doping distributions.

2. Effective Potential in Device Simulation

The effective potential concept uses the fact that as the electron moves, the edge of the wave packet encounters variations in the potential profile before the center of the wave packet. Mathematically, this effect at a point (x_i, y_j, z_k) can be treated as the convolution of the potential with a Gaussian wave packet to obtain the effective potential at this point as follows:

$$V_{\text{eff}} = \iiint V(x, y, z) G(x - x_i, y - y_j, z - z_k) dx dy dz$$

where G is a Gaussian function with a given standard deviation in each of the three coordinate directions. The spread of the wave packet can be determined by the thermal de Broglie wavelength for the lateral directions (Ferry *et al.* 2002, Ferry 2001). In the transverse direction (normal to the gate) it can be calculated based on the confining conditions (Ferry 2000, Ferry *et al.* 2002). In this work we use a value of 0.64 nm in the transverse direction, and 2.2 nm in the lateral directions.

We include the effective potential to treat the quantum effects in the SOI MOSFET structure depicted in

*Work supported by the Semiconductor Research Corporation.

[†]To whom correspondence should be addressed.

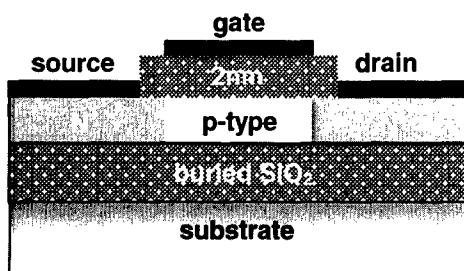


Figure 1. SOI MOSFET device structure used for all simulations.

Fig. 1. The SOI film thickness considered in this work was 5 nm, a channel length of 40 nm, and a device width of $0.48 \mu\text{m}$ was used to increase the number of electrons for the ensemble averages. The source and drain doping was at $2 \times 10^{19} \text{ cm}^{-3}$ and the channel doping and distribution varied for different simulations. The buried oxide thickness was 30 nm and the gate oxide thickness was 2 nm. The general procedure was to solve the Poisson equation first, then do the convolution with the Gaussian function in order to obtain the effective potential. It is the effective potential that is then used to calculate the electric fields that drift the electrons in the Monte Carlo transport kernel.

An example of an effective potential profile is shown in Fig. 2 for a device with channel doping of $5 \times 10^{18} \text{ cm}^{-3}$ and applied gate and drain voltages of 0.6 V. The effective potential steeply increases at the oxide interfaces as a result of the convolution with the Gaussian function representing the electron wave

packet. As a result of this potential increase, the electrons experience a strong electric field, which repels them from the interface. Therefore, the electrons are set-back from the gate oxide interface, resulting in a decrease in both gate capacitance and inversion charge density. Because of this decreased inversion charge density, a higher gate voltage is required to obtain the same inversion charge that would be present without the effective potential, which was described in Ramey and (Ferry 2001).

3. Random Doping Distribution Results

Previous simulations using the effective potential have focused on doping distributions that are uniformly distributed throughout the device regions. It is of interest therefore to determine how the effective potential simulations respond to a random discrete dopant distribution. A discrete distribution tends to create spikes in the potential profile, which behave like coulomb scattering centers. The effective potential, however, smoothes the potential profile, and thus alters the simulated interaction between the electron and the dopant ion and could therefore negate the effect of the discrete dopant ion.

To examine this effect, simulations were performed on devices with 5 nm silicon film thickness and various doping levels. Figure 3 illustrates the effect of the random dopant distribution for various doping levels in the channel. At the doping level of $5 \times 10^{17} \text{ cm}^{-3}$ there is virtually no difference in the I_D-V_G curves simulated with discrete and uniform doping distributions.

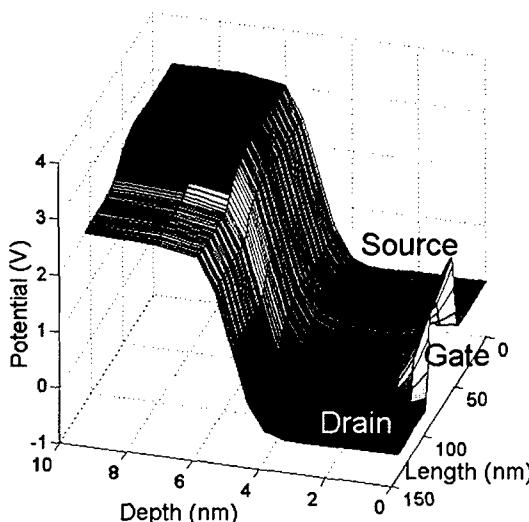


Figure 2. Effective potential profile for the SOI MOSFET device.

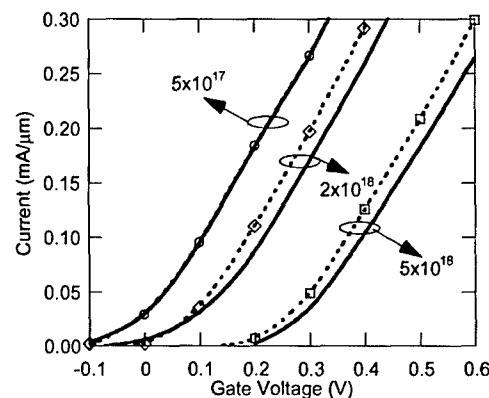


Figure 3. I_D-V_G curves for three different doping levels. The dashed lines with open symbols are for uniform doping, the solid lines for random doping distribution results averaged from several devices. All simulations were performed with the effective potential and $V_D = 0.1 \text{ V}$.

This can be easily understood simply by the number of dopant atoms in the channel of such a small device. At this doping level, there are only about 48 atoms in the channel, so there aren't enough to significantly affect the current. Simulations performed on devices at even lower channel doping reveal similar behavior. At higher doping levels, there is a noticeable shift in the I_D - V_G curves, with the discrete doping resulting in less current. As a result, the threshold voltage shifts about 25 mV higher for the simulations using discrete doping distributions.

The curves generated for the discrete doping devices are an average from several devices with different random doping distributions. There are certain configurations of dopant ions that can actually lead to lower threshold voltage than the uniform distribution. For example, if the dopants are located very close to the gate interface, their effect on the local electric field is essentially eliminated by the strong electric field formed by the effective potential profile at the interface. Conversely, if the dopants are situated far away from the interface, the inversion layer forms near the interface where there is a lower dopant density, and thus the threshold voltage is also reduced.

One possible explanation for this behavior would be that the electron density in the channel for the simulations with the random doping is lower than for simulations with uniform doping. However, Fig. 4 indicates that there are roughly equal amounts of carriers in the channel for each type of simulation. The plot does show slightly more variation in the sheet density

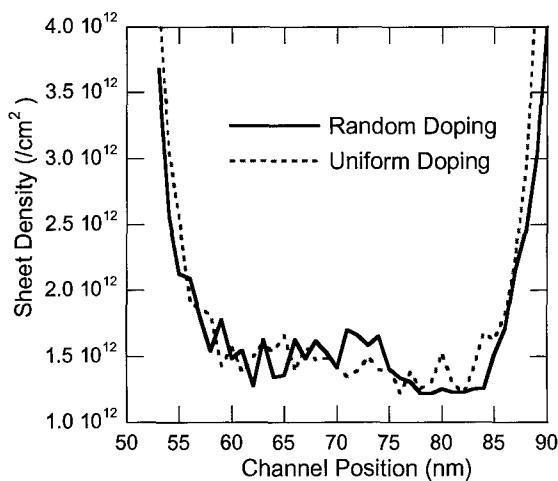


Figure 4. Sheet density in channel for the device with $N_A = 5 \times 10^{18} \text{ cm}^{-3}$ with random doping (solid-line) and uniform doping (dashed line). $V_G = 0.5 \text{ V}$, $V_D = 0.1 \text{ V}$.

for the random doping distribution, which is expected since there are regions of the channel with less dopant than in the uniform situation.

The situation can be explored qualitatively by examining the actual distribution of electrons in the channel. Figure 5 illustrates this distribution for simulations with discrete and uniform dopant distributions. The electron distribution in the channel for the uniform dopant distribution (top) is much smoother and more uniform than for the case of the discrete dopant distribution (bottom). Further, for the discrete dopant distribution simulation there are regions of very low electron density (at positions of about 62 and 78 nm) that effectively pinch off the channel. The pinched off regions correspond to high dopant concentration, and the region from 62 to 78 nm corresponds to a region of lower dopant concentration. As a result, a quantum dot effectively forms between 62 and 78 nm, since the discrete dopants form barriers that isolate a potential well. Such behavior is to be expected with random distributions, and has been discussed elsewhere (Ferry and Barker 1998).

The constriction indicated in the bottom panel of Fig. 5 gives a clear demonstration of the effect of the discrete dopant distribution. As a result of such constrictions and the random fluctuations in the potential profile, one would expect the velocity of the electrons

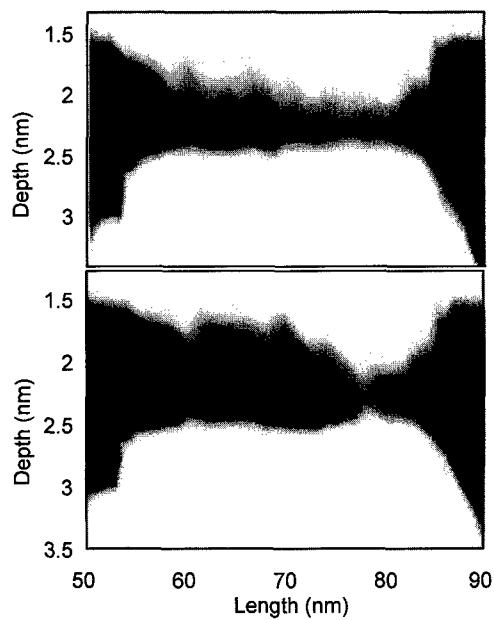


Figure 5. Electron distribution in channel for the device with $N_A = 5 \times 10^{18} \text{ cm}^{-3}$ and $V_G = 0.5 \text{ V}$. The top shows the uniform dopant distribution case, and the bottom the random distribution simulation. Darker regions indicate higher electron density.

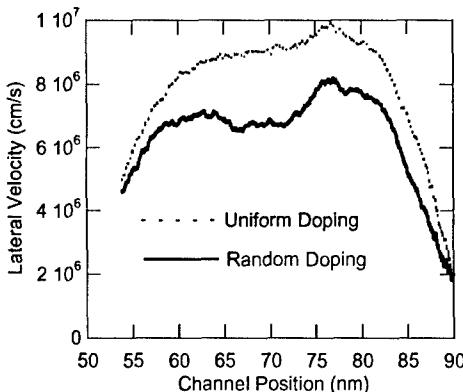


Figure 6. Lateral velocity of electrons in channel for $N_A = 5 \times 10^{18} \text{ cm}^{-3}$, $V_G = 0.5 \text{ V}$, and $V_D = 0.1 \text{ V}$. The random doping simulation is the solid-line and the uniform doping dashed line.

from source to drain to be lower as they scatter off the potential variations. This is indeed what happens, and is apparent from the large shift in the lateral velocity indicated in Fig. 6. Here, the lateral velocity is seen to be both significantly lower for the random distribution than the uniform distribution as well as have larger variation along the length of the channel. The larger variation of the velocity along the channel for the discrete distribution is due to the fact that the electron density varies from the random doping, as was indicated in Fig. 5. Therefore, the discrete doping is seen to cause a lower lateral velocity, which accounts for the lower current observed for the random doping simulations depicted in Fig. 3.

It is also interesting to compare the results of simulations with and without the effective potential for devices with discrete and uniform dopant distributions. To examine this, I_D - V_D curves were generated at a channel doping level of $2 \times 10^{18} \text{ cm}^{-3}$, which was seen previously to provide enough dopants to cause a measurable shift in the threshold voltage when the doping was discretely distributed. Figure 7 shows the results of these simulations, and as would be expected from the I_D - V_G behavior, the discrete doping results in a lower drain current than uniform doping when simulated with the effective potential. For simulations without the effective potential, there is a similar shift in the saturation current for the results with discrete and uniform doping. (Note: the simulations using random, discrete dopant distributions performed with and without the effective potential used the same discrete dopant distribution.) This indicates that the use of the effective potential does not significantly alter the interaction between the electrons and the ionized dopant atoms.

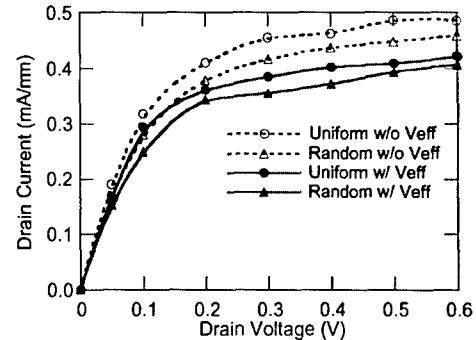


Figure 7. I_D - V_D curves for $N_A = 2 \times 10^{18} \text{ cm}^{-3}$ and $V_G = 0.4 \text{ V}$. The solid lines are for simulations with the effective potential, the dashed lines for simulations without the effective potential. The simulations with random doping are indicated by curves with triangles, and uniform doping by curves with circles.

As also can be seen in the figure, the simulations without the effective potential result in higher saturation current than simulations with the effective potential. This is due to the fact that the effective potential causes a charge set-back and elevates the ground state energy, which results in lower channel density. This is consistent with results explored in more detail previously for simulations with and without the effective potential using uniform doping distributions (Ramey and Ferry 2002). As clearly can be seen in Fig. 7, the use of the effective potential significantly affects the saturation current obtained, as does the use of a random doping distribution.

4. Conclusions

These simulations demonstrate the importance of including discrete random dopant distributions in ultra-small SOI MOSFET simulation. The use of random dopants tends to shift the calculated threshold voltage approximately 25 mV higher for these devices at heavy channel dopings, but has very little effect at lighter doping levels. The shift is due to a reduced carrier velocity in the channel that arises from scattering from the isolated dopant atoms. As a result, regions of high and low electron density develop, and structures resembling open quantum dots are observed.

The use of the effective potential in these types of simulations was a potential concern since the effective potential smoothes variations in the potential formed by the discrete dopant atoms. However, a similar shift in saturation current is obtained between simulations with discrete and uniform distributions, regardless of

whether the effective potential is employed. Therefore, the use of the effective potential approach is justified for simulations of discrete random doping distributions.

Acknowledgments

The support of the Semiconductor Research Corporation for this work is gratefully acknowledged. Also, the authors appreciate useful discussions with D. Vasileska and W. Gross.

References

- Ferry D.K. 2000. SuperLatt. Microstruct. 27: 61.
- Ferry D.K. 2001. VLSI Design 13: 155.
- Ferry D.K. and Barker J.R. 1998. VLSI Design 8: 165.
- Ferry D.K., Ramey S.M., Shifren L., and Akis R. J. Comp. Electron. 1: 59.
- Gross W.J., Vasileska D., and Ferry D.K. 2000. IEEE Trans. Elec. Dev. 47: 1831.
- Ramey S.M. and Ferry D.K. 2002. Physica B: Condensed Matter. 314: 350.
- Zhou J.R. and Ferry D.K. 1995. IEEE Computational Science and Engineering 2(2): 30.



Low-Field Mobility and Quantum Effects in Asymmetric Silicon-Based Field-Effect Devices

I. KNEZEVIC, D. VASILESKA, X. HE, D.K. SCHRODER AND D.K. FERRY

Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University,
Tempe, AZ 85287-5706, USA

irenak@asu.edu

Abstract. Though asymmetric MOSFET structures are being designed in response to small-geometry effects, the performance estimates of such devices often rely on the conventional device description, and neglect to properly account for the interplay between quantum effects and the effects of asymmetry. In this paper, we investigate the low-field transport in a highly asymmetric MOSFET structure, characterized by a p^+ -implant at the source end, by using a Monte Carlo—Poisson simulation with the quantum effects incorporated through an effective potential. We observe that highly-pronounced asymmetry leads to ballistic transport features, which become suppressed by the inclusion of quantum effects. We prove that mobility degradation is an essentially non-equilibrium signature of quantum mechanics, independent of the well-established equilibrium signatures (charge set-back and gap widening). Consequently, in order to properly estimate the device performance, it becomes important to account for the channel mobility degradation due to quantum effects.

Keywords: low-field transport, electron mobility, Monte Carlo simulation, asymmetric structures

1. Introduction

Asymmetrically doped metal-oxide-semiconductor field-effect transistors (MOSFETs) have recently received much attention due to the current quest to optimize the transistor performance simultaneously with its continuing shrinking, and overcome the inevitable accompanying increase in the severity of small-geometry effects. Representative asymmetric structures, which show improved performance with respect to some of the detrimental small-geometry effects, include the lightly-doped drain (LDD) devices, gate overlapped LDD structures (GOLD), halo source GOLD drain (HS-GOLD) (Buti *et al.* 1991), graded-channel MOSFETs (GCMOS) (Ma *et al.* 1997), and focused-ion-beam MOSFETs (FIBMOS) (Shen *et al.* 1998, Kang and Schroder 2000, Kang *et al.* in press). However, very often the performance of these devices is predicted according to models that hold for conventional MOSFETs, which are inadequate for several reasons. First, unlike conventional MOSFETs,

these devices are small and are therefore expected to experience quantum transport under bias. Secondly, the optimized doping profiles of such devices produce highly inhomogeneous electric fields, which may lead to non-stationary transport features, like velocity overshoot, even under steady-state conditions. The interplay between the small and the asymmetric has not been fully understood yet, but it certainly holds promise for some new and exciting transport phenomena.

In this paper, we present the results of a Monte Carlo particle-based simulation of low-field transport in an asymmetric MOSFET structure, with quantum effects included through an effective potential (Ferry 2000). The asymmetric structure simulated is characterized by a highly-doped ($1.6 \times 10^{18} \text{ cm}^{-3}$), narrow (70 nm) p^+ -implant, located near the source end of the 250 nm channel of a conventional MOSFET, with substrate doping equal to 10^{16} cm^{-3} (Fig. 1).

Such a structure could, for instance, be realized by using focused ion beam implantation (FIBMOS) (Shen *et al.* 1998, Kang and Schroder 2000, Kang *et al.* in

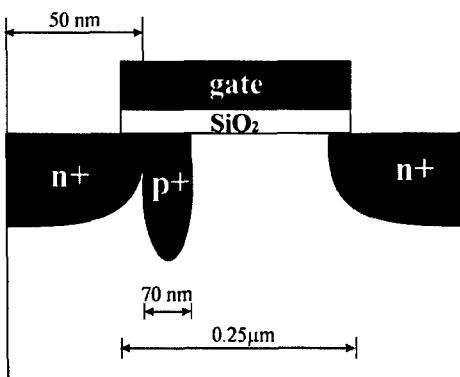


Figure 1. Schematic representation of the simulated asymmetric MOSFET structure.

press). There are several reasons for choosing such a structure: it is fairly easy to simulate, due to the simple geometry; the asymmetry effects should be observable because of the very abrupt changes in the doping profile, and the highly-doped implant region promises quantum effects even if the device as a whole is large.

The quantum effects in this study have been accounted for by including an effective potential (Ferry 2000) in the classical particle simulator. This has proven quite successful in treating the one-particle quantum effects in inversion layers (Ferry *et al.* 2000, Knezevic *et al.* 2002, Ramey and Ferry 2002). First, we will present the transfer characteristics of the device with and without the effective potential, and point out the main features that quantization introduces at the macro level of device analysis. We will then briefly review the microscopic quantum transport effects that have been known to contribute to the output trends obtained (Ferry *et al.* 2000, Knezevic *et al.* 2002, Ramey and Ferry 2002), and then focus on what has not received sufficient attention so far, and that is the behavior of low-field mobility when quantum effects are included in a highly asymmetric structure. The dependence of low-field mobility on the lateral electric field will be analyzed, with respect to both the introduction of asymmetry and quantum effects.

2. Macroscopic Signatures of Quantum Effects

The transfer characteristic of the simulated device structure with and without the inclusion of quantum effects, for the drain voltage $V_D = 0.4$ V, is shown in Fig. 2. It is clear that the threshold voltage, V_{th} , is higher with the inclusion of the effective potential. Also, the device transconductance, $g_m = dI_D/dV_G$, in the linear

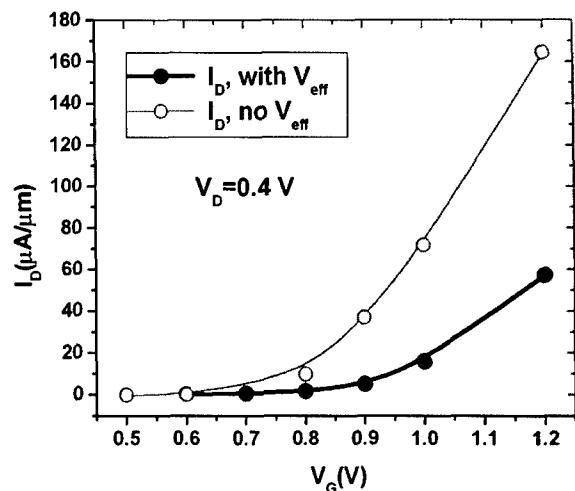


Figure 2. Transfer characteristic of the simulated asymmetric device structure, for drain bias $V_D = 0.4$ V, with and without the inclusion of quantum effects through V_{eff} .

region is clearly lower if the effective potential, V_{eff} , is included.

It has been shown (Ferry *et al.* 2000, Knezevic *et al.* 2002, Ramey and Ferry 2002) that the inclusion of quantum effects in the microscopic description of the inversion layer of a metal-oxide-semiconductor device leads to two major features: reduced sheet density of channel carriers and charge set-back from the semiconductor/oxide interface. The reduced sheet density leads to an increase in the threshold voltage and a decrease in the drive current. On the other hand, the charge set-back leads to an effective increase in the oxide thickness, thereby degrading the device transconductance. Even though these two microscopic features undoubtedly have a very important impact on the transfer characteristics presented in Fig. 2, that may not be the entire story. Namely, there is very little change in the sheet density and the charge set-back between the equilibrium conditions and the non-zero drain bias situation, which means that these are virtually equilibrium quantum-mechanical effects. However, if the device is on and a bias is applied between the source and the drain, especially with an asymmetric doping profile such as that of the simulated device, some purely non-equilibrium features emerge.

3. Low-Field Mobility and the Interplay of Quantum Mechanics and Asymmetry

According to Fig. 2, the bias condition $V_G = 1.2$ V with V_{eff} is almost equivalent to $V_G = 1.0$ V without V_{eff} :

the current is similar, and $V_G - V_{th}$ is virtually the same. Similarity is also noted between $V_G = 1.0$ V with V_{eff} and $V_G = 0.9$ V without V_{eff} . According to Fig. 3(a), which presents the profile of the effective perpendicular field in the channel for the above bias conditions, we indeed note that the equilibrium part of the quantum-mechanical influence, as described in the previous paragraph, is the same for $V_G = 1.2$ V with and $V_G = 1.0$ V without V_{eff} , and for $V_G = 1.0$ V with and $V_G = 0.9$ V without V_{eff} , so throughout this paper we will deal with these two pairs of bias conditions, as they will help us isolate the non-equilibrium signatures of quantum-mechanical effects.

By the effective field we mean the field felt by electrons, which is found according to

$$E_{eff}(x) = \frac{\int_0^{y_{max}} E(x, y) n(x, y) dy}{\int_0^{y_{max}} n(x, y) dy}. \quad (1)$$

In (1), n is the electron concentration, the x -coordinate runs along the channel, starting at the source, and the y -axis is perpendicular to the semiconductor/oxide interface, starting at the interface and ending at the device bottom boundary (y_{max}). This definition is meaningful only in inversion ($n > 0$). (The effective field has nothing to do with the effective potential, despite the names.) Figure 3(b) shows the profile of the lateral effective field, while Fig. 3(c) presents the profile of the average carrier velocity in the channel. Note that the initially noisy raw Monte-Carlo output profiles have been smoothed.

The inclusion of the effective potential apparently leads to lower average velocity. Due to the pronounced asymmetry, the velocity overshoot is evident when the electrons just exit the implant and are subject to large negative electric fields. The inclusion of V_{eff} suppresses the overshoot. Also, at about 200–250 nm, we note that the lateral field is very low or even zero, whereas the velocity is finite, which signalizes that non-local effects are important (velocity is not correlated with the field at the given point), and transport can hardly be regarded as diffusive.

If the lateral field is low (and negative), below 10 kV/cm in magnitude, we can speak about low-field transport, and the low-field mobility profile along the channel can be found as

$$\mu(x) = \frac{v(x)}{E_{x,eff}(x)}. \quad (2)$$

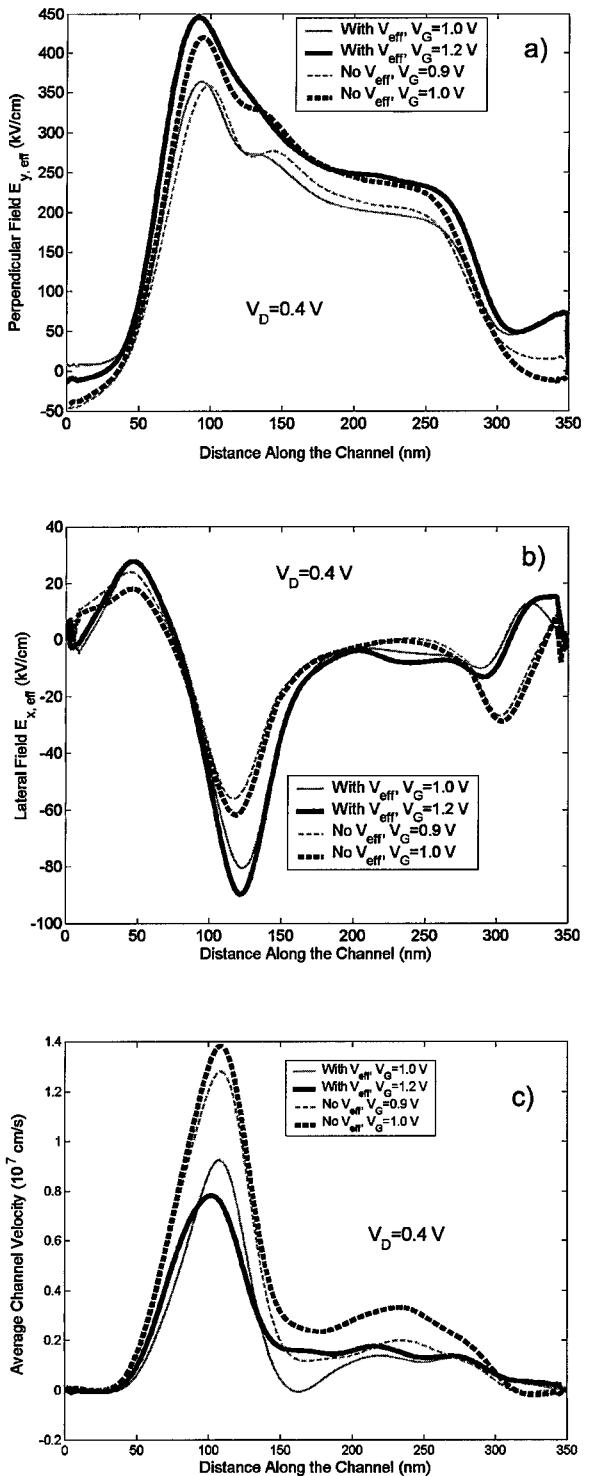


Figure 3. Profiles of the (a) perpendicular effective electric field, (b) lateral effective electric field and (c) average carrier velocity in the channel, with and without V_{eff} , for various gate bias conditions, and $V_D = 0.4$ V.

The profiles $E_{x,\text{eff}}(x)$, $E_{y,\text{eff}}(x)$, $\mu(x)$ obtained this way for a given V_D , V_G , actually represent a parametric equation of a curve on the surface $\mu(E_{x,\text{eff}}, E_{y,\text{eff}})$. Even if we had infinitely many different bias conditions V_D , V_G , we could never completely reconstruct the entire surface. However, we note that between 130 nm and 300 nm the perpendicular field is fairly monotonic, and if we restrict ourselves to the areas where the lateral field is monotonic as well, we are guaranteed that the curve $(E_{x,\text{eff}}(x), E_{y,\text{eff}}(x), \mu(x))$ on the $\mu(E_{x,\text{eff}}, E_{y,\text{eff}})$ surface gives single-valued projections onto $E_{x,\text{eff}} = \text{const.}$ or $E_{y,\text{eff}} = \text{const.}$ planes. After plotting $(E_{x,\text{eff}}(x), E_{y,\text{eff}}(x), \mu(x))$ for regions between 130 nm and 300 nm with monotonic $E_{x,\text{eff}}(x)$, we find that the curve shows virtually no dependence on the perpendicular field within a given range of lateral fields. This is not surprising, as our simulation does not include surface roughness scattering, but only acoustic and intervalley phonon scattering. Therefore, we may conclude that

$$\mu(E_{x,\text{eff}}, E_{y,\text{eff}}) \approx \mu(E_{x,\text{eff}}(x)) \approx (\mu(E_{x,\text{eff}})) \text{ averaged over all } E_{y,\text{eff}} \text{ in a given range} \quad (3)$$

Figure 4 shows the low-field mobility as a function of the lateral field, for several values of the average perpendicular field as a parameter. As the lateral field approaches zero, the mobility increases, both with and without V_{eff} , which signalizes ballistic transport (in other words, non-locality; the retardation effects become important). However, the quantum mechanical behavior suppresses the ballistic feature. Even at

somewhat higher fields, the decrease in mobility due to the inclusion of V_{eff} is significant, which indicates that it is important to include the influence of quantum mechanical effects on mobility in order to have a realistic physical picture of device operation.

4. Conclusions

In this paper we have investigated the influence of quantum-mechanical effects on low-field transport in a highly asymmetric MOSFET structure (Fig. 1). By analyzing the behavior with and without V_{eff} , we identified pairs of gate biases, such that one voltage in the pair corresponds to V_{eff} included and the other to V_{eff} excluded, and both lead to equivalent inversion conditions. Comparing between the two biases in the pair enables us to set aside the essentially equilibrium quantum mechanical effects, and just observe the non-equilibrium role of quantization in the simulated structure.

There are several important conclusions to be drawn. First, pronounced asymmetry leads to ballistic transport features. This feature is suppressed if the quantum-mechanical effects are included. Highly asymmetric devices therefore show signatures of ballistic transport, as seen also in the behavior of low field mobility. Even though quantum effects do lower the mobility, there is a definite tendency of mobility increase as the lateral field decreases, both with and without the effective potential. Altogether, asymmetry may lead to faster devices, but whether transport is truly ballistic or still diffusive needs to be carefully assessed by including quantum-mechanical effects into mobility modeling.

Acknowledgments

The authors would like to thank Massimo M. Fischetti, Steven M. Goodnick, Srdjan Milicic, Salvador Gonzalez and Gil Speyer for valuable discussions. Financial support from the Semiconductor Research Corporation, the Office of Naval Research under Contract No. N000149910318 and the National Science Foundation under NSF-CAREER ECS-9875051 is gratefully acknowledged.

References

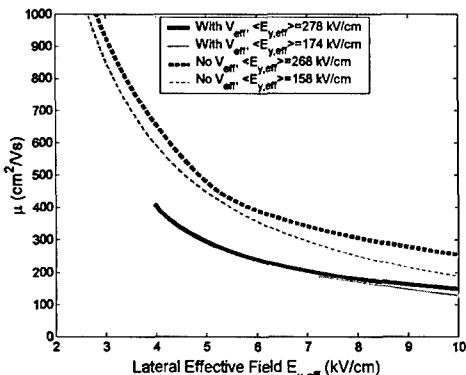


Figure 4. Variation of low-field mobility with the lateral effective field, with and without V_{eff} , for several perpendicular electric fields.

- Buti T.N. et al. 1991. IEEE Trans. Electron Dev. 38: 1757.
Ferry D.K. 2000. Superlattices Microstruct. 27: 61.

- Ferry D.K. *et al.* 2000. IEDM Tech. Dig. 287.
- Kang J. and Schroder D.K. 2000. In: Tech. Proc. of the Third International Conference on Modeling and Simulation of Microsystems, San Diego, California, March 27–29. p. 356.
- Kang J. *et al.*, in press.
- Knezevic I. *et al.* 2002. IEEE Trans. Electron Dev. 49: 1019.
- Ma J. *et al.* 1997. IEEE Trans. Very Large Scale (VLSI) Syst. 5: 352.
- Ramey S.M. and Ferry D.K. 2002. Physica B 314: 350.
- Shen C.C. *et al.* 1998. IEEE Trans. Electron Dev. 45: 453.



Quantum Potential Corrections for Spatially Dependent Effective Masses with Application to Charge Confinement at Heterostructure Interfaces

J.R. WATLING, J.R. BARKER AND S. ROY

Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow, G12 8LT, UK

J.Barker@elec.gla.ac.uk

Abstract. The effect of a spatially varying effective mass as encountered in heterostructure devices is shown to lead to classical and quantum corrections to the description of transport. The quantum potential corrections for pure states and the corrections to Quantum Monte Carlo for mixed states are derived using the Wigner formalism. The application to SiGe graded structures is shown to lead to additional corrections, which are of the same order as the conventional density gradient corrections.

Keywords: Density gradient, Wigner function, semiconductor devices

1. Introduction

With the advent of decanano scale semiconductor devices it has become important to include quantum corrections to conventional device modelling in a pragmatic fashion. The density gradient or quantum potential (Bohm 1952a, b) has been deployed within hydrodynamic and drift-diffusion modelling. More recently, expansions (Ancona and Iafrate 1989, Tsuchiya and Miyoshi 2000, Tsuchiya, Fischer and Hess 2000) of the non-local Wigner equation of motion to second order in Planck's constant have led to so-called Quantum Monte Carlo models. However, to date, none of these formalisms have been consistent with the presence of a heterojunction with a spatially dependent effective mass. In the present paper we show that additional quantum corrections are required to incorporate situations where the effective mass varies with position as it does for transport in heterostructures and in particular for transport in devices based upon $Si_{1-x}Ge_x$ heterostructure devices. For a purely classical model the simple Hamiltonian $H = \mathbf{p}^2/2m(\mathbf{r}) + V(\mathbf{r})$ generates an effective force due to a varying mass given by $F_{eff} = (\mathbf{p}^2/2)\nabla m^{-1}(\mathbf{r}) = (\mathbf{p}^2/2m(\mathbf{r}))\nabla \ln m(\mathbf{r})$. For a mass discontinuity at an interface in an idealised heterostructure the classical effective force due to the mass

change is impulsive:

$$F_{eff} = -\frac{p^2}{2} \left(\frac{1}{m_2^*} - \frac{1}{m_1^*} \right) \delta(x - x_{interface}) \quad (1)$$

2. Pure State Quantum Potential Corrections

In the case of a pure state, $\Psi(\mathbf{x}, t) = \langle \mathbf{x} | \Psi(t) \rangle = R(\mathbf{x}, t) \exp[\frac{i}{\hbar} S(\mathbf{x}, t)]$, written here in polar form, a quantum potential V_Q may be obtained directly from the effective mass Schrödinger's equation using the construction

$$V_Q = \frac{Re \langle \Psi | \mathbf{x} \rangle \hat{H}_0(\hat{\mathbf{x}}, \hat{\mathbf{p}}) \langle \mathbf{x} | \Psi \rangle}{|\langle \mathbf{x} | \Psi \rangle|^2} - H_0(\mathbf{x}, \nabla S) \quad (2)$$

where H_0 is the effective mass Hamiltonian.

If we consider a *minimal* hermitian Hamiltonian, taking into account a position-dependent effective mass tensor in the form of the BenDaniel-Duke Hamiltonian (BenDaniel and Duke 1966):

$$H = H_0 + V(\mathbf{r}) \quad H_0 = \mathbf{p} \frac{1}{2m(\mathbf{r})} \mathbf{p} \quad (3)$$

the time-dependent Schrödinger equation may be written as:

$$-i\hbar \frac{\partial \Psi}{\partial t} = H\Psi = -\frac{\hbar^2}{2}\nabla \left(\frac{1}{m^*} \nabla \Psi \right) + V\Psi \quad (4)$$

This leads to a new quantum potential, $V_{Qhetero}$, of the form:

$$V_{Qhetero} = \frac{-\hbar^2}{2m^*} \frac{\nabla^2 |\Psi|}{|\Psi|} + \frac{\hbar^2}{2} \frac{\nabla m^*}{m^{*2}} \frac{\nabla |\Psi|}{|\Psi|} \quad (5)$$

where the carrier density, n , may be interpreted as $n \propto |\Psi|^2$. Thus, the quantum potential in the presence of a heterointerface includes an additional term that is dependent on the gradient of the effective mass. The importance of this new, previously neglected, term can clearly be seen in Figs. 1 and 2 where we display the wavefunctions and corresponding contributions to the quantum potential for a 7.5 nm Si/Si_{0.5}Ge_{0.5}/Si quantum well with graded interfaces.

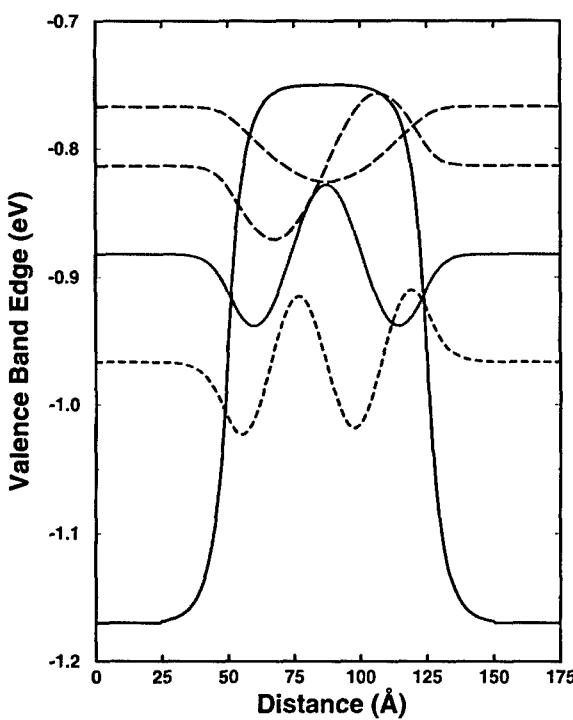


Figure 1. First four eigenfunctions for a 75 Å Si/Si_{0.5}Ge_{0.5}/Si quantum well, with 'soft' interfaces.

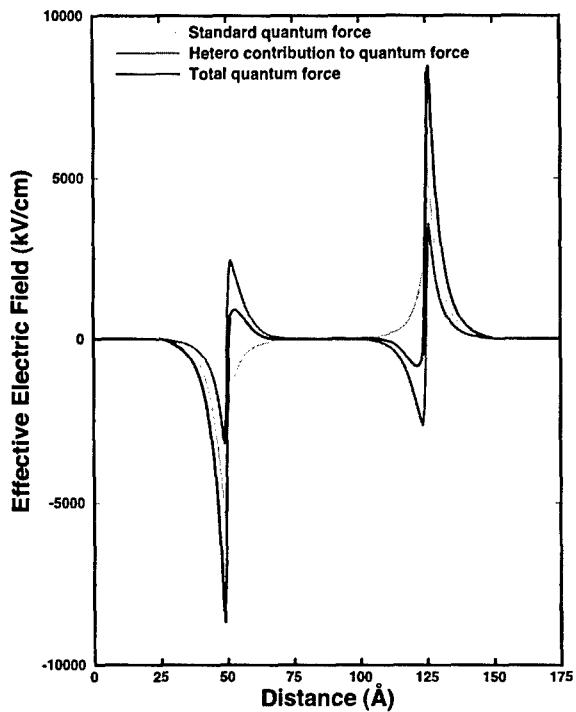


Figure 2. The total quantum force arising from the first eigenstate shown in Fig. 1.

3. Mixed State Quantum Monte Carlo Corrections

The simple *pure state* quantum potential is less helpful for modelling quantum corrections in a self-consistent time-dependent potential $V(\mathbf{r})$. Thus for device modelling one is more generally interested in *mixed state* quantum transport formalisms typified by the Wigner or density matrix equations of motion. The first study of space-dependent effective mass corrections to the Wigner equation were derived by Barker, Lowe and Murray (1984) using the Ben Daniel-Duke Hamiltonian to give an exact result which showed that the Wigner equation of motion comprises driving terms which are integrals over phase of the non-local effective force and derivatives of the Wigner function $f(\mathbf{r}, \mathbf{p}, t)$ weighted by Si and Ci function kernels. However this study (Bohm 1952a, b) did not examine the local approximations to order \hbar^2 , which have recently come in vogue for quantum Monte Carlo (Tsuchiya and Miyoshi 2000, Tsuchiya, Fischer and Hess 2000).

The exact Wigner equation may be derived as the Wigner transform of the density matrix equation using the basic theorem: the product AB of two operators A, B with corresponding Wigner Transforms $A(\mathbf{r}, \mathbf{p})$,

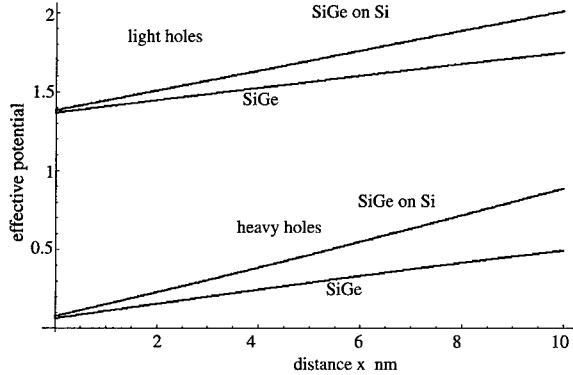


Figure 3. Variation of effective potential $\beta V_{\text{mass}} = -\ln m^*$ for a linearly graded region in the range $c = [0, 0.3]$, $d = 10 \text{ nm}$.

$B(\mathbf{r}, \mathbf{p})$ is

$$C(\mathbf{r}, \mathbf{p}) = \exp \left[\frac{i\hbar}{2} (\nabla_{\mathbf{p}}^B \cdot \nabla_{\mathbf{r}}^A - \nabla_{\mathbf{r}}^B \cdot \nabla_{\mathbf{p}}^A) \right] \times A(\mathbf{r}, \mathbf{p}) B(\mathbf{r}, \mathbf{p}) \quad (6)$$

Using (6) the Wigner transform of the variable mass Hamiltonian is:

$$H_0(\mathbf{r}, \mathbf{p}) = \frac{p^2}{2m(\mathbf{r})} + \frac{\hbar^2}{8} \nabla^2 \frac{1}{m(\mathbf{r})} \quad (7)$$

The Wigner transform of the density matrix equation of motion then gives the corresponding Wigner equation for varying effective mass as:

$$\frac{\partial f}{\partial t} - \frac{2}{\hbar} \sin \left[\frac{\hbar}{2} (\nabla_{\mathbf{p}}^f \cdot \nabla_{\mathbf{r}}^H - \nabla_{\mathbf{r}}^f \cdot \nabla_{\mathbf{p}}^H) \right] H(\mathbf{r}, \mathbf{p}) \times f(\mathbf{r}, \mathbf{p}, t) = 0 \quad (8)$$

$$H(\mathbf{r}, \mathbf{p}) = H_0(\mathbf{r}, \mathbf{p}) + V(\mathbf{r})$$

Expanding the expression (8) to $O(\hbar^2)$ we obtain the generalised Wigner equation in the density-gradient approximation:

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{p_{\mu}}{m} \frac{\partial f}{\partial x_{\mu}} - \frac{\partial V}{\partial x_{\mu}} \frac{\partial f}{\partial p_{\mu}} - \frac{p^2}{2} \frac{\partial m^{-1}}{\partial x_{\mu}} \frac{\partial f}{\partial p_{\mu}} + \frac{\hbar^2}{24} \\ \times \frac{\partial^3 V}{\partial x_{\mu} \partial x_{\nu} \partial x_{\kappa}} \frac{\partial^3 f}{\partial p_{\mu} \partial p_{\nu} \partial p_{\kappa}} + \frac{\hbar^2}{8} \\ \times \frac{\partial^3 m^{-1}}{\partial x_{\kappa} \partial x_{\mu} \partial x_{\nu}} \frac{\partial f}{\partial p_{\mu}} - \frac{\hbar^2}{4} \left\{ \frac{\partial m^{-1}}{\partial x_{\kappa}} \frac{\partial^3 f}{\partial x_{\mu} \partial x_{\nu} \partial p_{\kappa}} \right. \\ \left. - p_{\kappa} \frac{\partial^2 m^{-1}}{\partial x_{\mu} \partial x_{\nu}} \frac{\partial^3 f}{\partial p_{\mu} \partial p_{\nu} \partial x_{\kappa}} \right\} \\ + \frac{p^2}{6} \frac{\partial^3 m^{-1}}{\partial x_{\mu} \partial x_{\nu} \partial x_{\kappa}} \frac{\partial^3 f}{\partial p_{\mu} \partial p_{\nu} \partial p_{\kappa}} = 0 \end{aligned} \quad (9)$$

The first three terms of (9) are identical to the driving terms of a constant-mass Boltzmann equation. The fourth term is the classical correction due to the varying effective mass. The fifth term is the well-known quantum correction due to the potential V . The remaining terms are the quantum corrections due to the varying effective mass. The *non-locality* of the exact Wigner equation is here reflected in the presence of the higher derivatives of the Wigner distribution.

Following the approach of Tsuchiya and Miyoshi (2000) and Tsuchiya, Fischer and Hess (2000) we can easily recover the second order *local* quantum approximation to the effective force produced by the variable mass. The approach eliminates the higher derivatives of f by the ansatz: $f \approx \exp[-\beta(\mathbf{p} - \mathbf{p}_d)^2/2m - \beta V(r) - \beta \mu]$ to obtain the local quantum kinetic equation which underpins Quantum Monte Carlo simulation but with corrections due to the variable effective mass:

$$\frac{\partial f}{\partial t} + \frac{p_{\mu}}{m} \frac{\partial f}{\partial x_{\mu}} + (F_{\mu} + F_{Q\mu} + F_{Qm\mu}) \frac{\partial f}{\partial p_{\mu}} = 0 \quad (10)$$

where the classical force F includes the varying mass contribution F_{eff} , the quantum force F_Q derives from the potential V as

$$F_{Q\mu} = -\frac{\beta \hbar^2}{24m} \left[3 - \beta \frac{(\mathbf{p} - \mathbf{p}_d)^2}{m} \right] \frac{\partial^3 V}{\partial x_{\mu} \partial x_{\nu} \partial x_{\kappa}} \quad (11)$$

and finally the varying mass contributes a quantum force correction, which we only display for the *slowly varying mass approximation* that neglects terms of order $(\nabla m^{-1})^2$ and retains only the lowest derivatives of the reciprocal effective

$$F_{Qm\mu} = \left(\frac{-\hbar^2}{4} \right) \left\{ \frac{\partial m^{-1}}{\partial x_{\mu}} \beta \left[-\frac{\partial^2 V}{\partial x_{\kappa}^2} + \beta \left(\frac{\partial V}{\partial x_{\kappa}} \right)^2 \right] \right\} \quad (12)$$

The total correction to the constant mass driving force due to varying mass is thus:

$$\begin{aligned} F_{\text{mass}\mu} = & \left(\frac{1}{m} \frac{\partial m}{\partial x_{\mu}} \right) \\ & \times \left(\frac{p^2}{2m} + \beta \frac{\hbar^2}{4m} \left[\frac{\partial^2 V}{\partial x_{\kappa}^2} - \beta \left(\frac{\partial V}{\partial x_{\kappa}} \right)^2 \right] \right) \end{aligned} \quad (13)$$

By inspection it is seen that the scale of the quantum potential determines the quantum correction to the classical force:

$$|\Delta F/F| \approx \frac{1}{2} \frac{V_Q}{p^2/2m} \quad (14)$$

To estimate the size of the effect we represent the classical correction in terms of an effective potential V_{mass} ; for kinetic energy of the order of $k_B T$, we have in dimensionless form:

$$\beta V_{mass} = -\ln m^* \quad (15)$$

4. Applications

In heterostructures such as $\text{Si}_{1-c}\text{Ge}_c$, we often encounter a linearly graded change in concentration c over a distance $L : c = c^*x/d$, where c^* is the final concentration. The density of states effective mass ratios for heavy holes in $\text{Si}_{1-c}\text{Ge}_c$ and $\text{Si}_{1-c}\text{Ge}_c$ on Si are given respectively by:

$$\begin{aligned} m_A^* &= 0.94 - 1.44c + 1.146c^2; \\ m_B^* &= 0.927 - 2.266c + 1.827c^2 \end{aligned} \quad (16)$$

The above masses were obtained by fitting to the density of states masses from a 6-band $\mathbf{k} \cdot \mathbf{p}$ calculated

over the range $c=0$ to $c=0.4$. In Fig. 3 we plot the results of our calculations for BVmass, for the situation described above.

5. Conclusions

We have demonstrated the quantum Monte Carlo force corrections required for the slowly varying mass approximation. The results indicate that the corrections are of the same order as the conventional quantum potential corrections in silicon-germanium systems with linear grading.

References

- Ancona M.G. and Iafrate G.J. 1989. Physical Review B 35: 9536–9540.
- Barker J.R., Lowe D.W., and Murray S. 1984. The Physics of Sub-micron Structures. pp. 277–286.
- BenDaniel D.J. and Duke C.B. 1966. Physical Review 126: 1386–1393.
- Bohm D. 1952a. Physical Review 85: 166–179.
- Bohm D. 1952b. Physical Review 85: 180–193.
- Tsuchiya H., Fischer B., and Hess K. 2000. IEDM Tech. Digest. pp. 283–286.
- Tsuchiya H. and Miyoshi T. 2000. Superlattices and Microstructures 27: 529–532.



Comparison of Three Quantum Correction Models for the Charge Density in MOS Inversion Layers

XINLIN WANG* AND TING-WEI TANG

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003, USA

xinlinw@us.ibm.com

Abstract. In order to obtain high density integration for MOS devices, it is necessary to reduce the gate oxide thickness and increase the substrate doping concentration. This results in a narrow and deep potential well in which electrons are confined at the semiconductor-insulator interface and it becomes necessary to take quantum mechanical (QM) effects into consideration. In this study, we compare three well established quantum correction models, i.e., the Hänsch model (Hänsch W. *et al.* 1989. Solid State Electronics 32(10): 839–849), the modified local density approximation (MLDA) model (Paasch G. and Ubensee H. 1982. Phys. Stat. Sol. (b) 113: 165–178), and the density-gradient (D-G) model (Ancona M.G. and Tiersten H.F. 1987. Physical Review B 35(15): 7959–7965; Ancona M.G. 1997. JTCAD 97–100) in terms of accuracy for predicting the inversion layer charge distribution.

Keywords: quantum mechanical effect, charge distribution, modified local density approximation, density-gradient theory

1. Introduction

When the quantum effect becomes noticeable in the deep-submicron MOSFETs, the Schrödinger-Poisson (S-P) equation is the most accurate way to handle the problem of the inversion-layer charge density, but it is not suitable for engineering applications especially for the two- and three-dimensional cases. Thus it is important to find a method which can produce a result similar to the quantum mechanically calculated one but requires only about the same computation cost as that of the classical calculation. In this work, different methods of quantum correction to the inversion layer charge density calculation have been studied. Calculations are carried out for 1-D Polycrystalline-Insulator-Semiconductor MOS structure with ⟨100⟩ oriented p-type silicon as substrate. No penetration of the wave function into the oxide is assumed. The carrier concentration for poly gate is $5 \times 10^{19} \text{ cm}^{-3}$, the oxide thickness t_{ox} is 3 nm and different doping profiles are used for

the silicon layer. The Fermi-Dirac distribution and the standard effective-mass approximation in a parabolic band are assumed. The parameter m_k appearing in the models has been determined by calibrating with the Schrödinger-Poisson (S-P) solutions.

2. Physical Fundamentals for Three Quantum Correction Models

Hänsch *et al.* (1989) gave the expression of the electron concentration with the QM correction as

$$n_{QM}(x) = N_C \exp\left(-\frac{q\psi(x) - q\varepsilon_F}{k_B T}\right) \times \left[1 - \exp\left(-\frac{x^2}{\lambda_{th}^2}\right)\right],$$
$$\lambda_{th} = \left(\frac{\hbar^2}{2m_n^* k_B T}\right)^{\frac{1}{2}}, \quad (1)$$

where λ_{th} is the thermal wavelength, m_n^* is the effective electron mass ($m_n^* = m_k \times 9.11 \times 10^{-31} \text{ kg}$), m_k is an adjustable parameter, x represents the distance from

*Present address: IBM, SRDC, 2070 Rte. 52, Hopewell Junction, NY 12533, USA.

the Si/SiO₂ interface, N_c is the conduction band effective density-of-states and ε_F is the Fermi level. This model gives an explicit expression for n_{QM} and therefore it is easy to be included in Poisson's equation by simply using n_{QM} instead of classical n .

Paasch and Ubensee (1982) first proposed the MLDA model and extended this method to the case where the potential has a large and abrupt change at a certain plane. When the QM correction is incorporated, the electron density near the Si/SiO₂ interface can be approximated by Jeong, Logan and Slinkman (1998)

$$\begin{aligned} n_{QM}(x) &= N_c \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{d\xi \cdot \xi^{\frac{1}{2}}}{1 + e^{\xi - k(x)}} \\ &\times \left\{ 1 - \sum_{i=1}^6 j_0(2x\sqrt{\xi}/\lambda_n^i)/6 \right\}, \\ k(x) &= \frac{\varepsilon_F + q\psi(x)}{k_B T}, \end{aligned} \quad (2)$$

where j_0 is the zeroth-order spherical Bessel function and λ_{th} is as defined in the Hänsch model.

The D-G model advanced by Ancona and his coworkers is an approximate approach to the QM correction of the macroscopic electron transport equation. In this approach, an extra term is introduced in the carrier flux by making the equation of state for the electron gas density-gradient dependent (Ancona and Tiersten 1987, Ancona 1997), i.e.,

$$\vec{J}_n = -qn\mu_n \nabla \psi + qD_n \nabla n - qn\mu_n \nabla \left(2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} \right) \quad (3)$$

where $b_n = \hbar^2/(12qm_n^*)$ is the (linear) density-gradient coefficient. In this study, b_n is treated as a fitting parameter. The D-G theory-based equation in 1-D under the non-tunneling condition is given by

$$2b_n \frac{d^2 y}{dx^2} + [\psi - \phi_F - T(y)]y = 0, \quad y \equiv \sqrt{n}, \quad (4)$$

where $T(y) = F_{1/2}^{-1}(y^2/N_C)$ for the Fermi-Dirac statistics; $T(y) = \ln(y^2/N_C)$ for the Boltzmann statistics. We apply a mixed discretization scheme (Wang 2001) as following. For any node $j < i$, f_{nj} is discretized linearly; and for the nodes $j > i$, the non-linear discretization scheme (Ancona and Biegel 2000) is applied. The node i designates the approximate position of the boundary layer. Assuming the Boltzmann statistics is valid and that $[\psi_i - \phi_F - 2\ln(y_i)]$ is a constant

within the integration interval, we have

$$\begin{cases} y = y_i - \frac{y_i - y_{i-1}}{h_i} h, & h = x - x_{i-1}, \quad x \in [x_{i-1}, x_i] \\ y = y_i \left(\frac{y_{i+1}}{y_i} \right) e^{\frac{h}{h_{i+1}}}, & h = x - x_i, \quad x \in [x_i, x_{i+1}] \end{cases} \quad (5)$$

and

$$\begin{aligned} f_{ni} &= \left[\frac{1}{h_{i+1}h_i} P \left(\frac{y_{i+1}}{y_i} \right) (y_{i+1} - y_i) - \frac{y_i - y_{i-1}}{h_i^2} \right] \\ &+ [\psi_i - \phi_F - T(y_i)] \\ &\times \left[\frac{3}{4}y_i + \frac{1}{4}y_{i-1} + y_i \frac{h_{i+1}}{h_i} 2Q \left(\frac{y_{i+1}}{y_i} \right) \right], \end{aligned} \quad (6)$$

where

$$\begin{aligned} P(z) &= \sqrt{z}B(\ln(z)), \quad Q(z) = \frac{1}{2B(\ln(\sqrt{z}))}, \\ B(z) &= \frac{z}{\exp(z) - 1}. \end{aligned}$$

It is also found that the electron concentration in the boundary layer (≤ 5 Å) under the inversion condition is proportional to the square of the distance to the Si/SiO₂ interface, i.e., $n \propto x^2$ (Wang 2001). Based on this fact, we implemented this mixed discretization scheme in which the linear scheme is used in the boundary layer and the nonlinear scheme is used in the rest of the silicon region.

3. Simulation Results

The poly-oxide-silicon MOS system is solved by 1-D S-P equation and other three theoretical quantum correction models described in the previous section. The solution obtained by the eight-subband S-P approximation is considered as the "exact" solution and used as a reference for comparison. The comparison criterion is based on the average displacement (the first moment) of charge distribution away from the oxide-silicon interface. Three types of MOS structures, which differ by their doping profiles in the silicon layer, are considered. The first structure has a uniform doping profile $N_a = 1 \times 10^{17} \text{ cm}^{-3}$. For the second one, we assume a low-high (retrograde) step doping profile, with surface doping $N_s = 1 \times 10^{17} \text{ cm}^{-3}$ and abruptly rising to $N_{bulk} = 1 \times 10^{18} \text{ cm}^{-3}$ at a 10 nm depth from the interface. The last one assumes a Gaussian doping profile

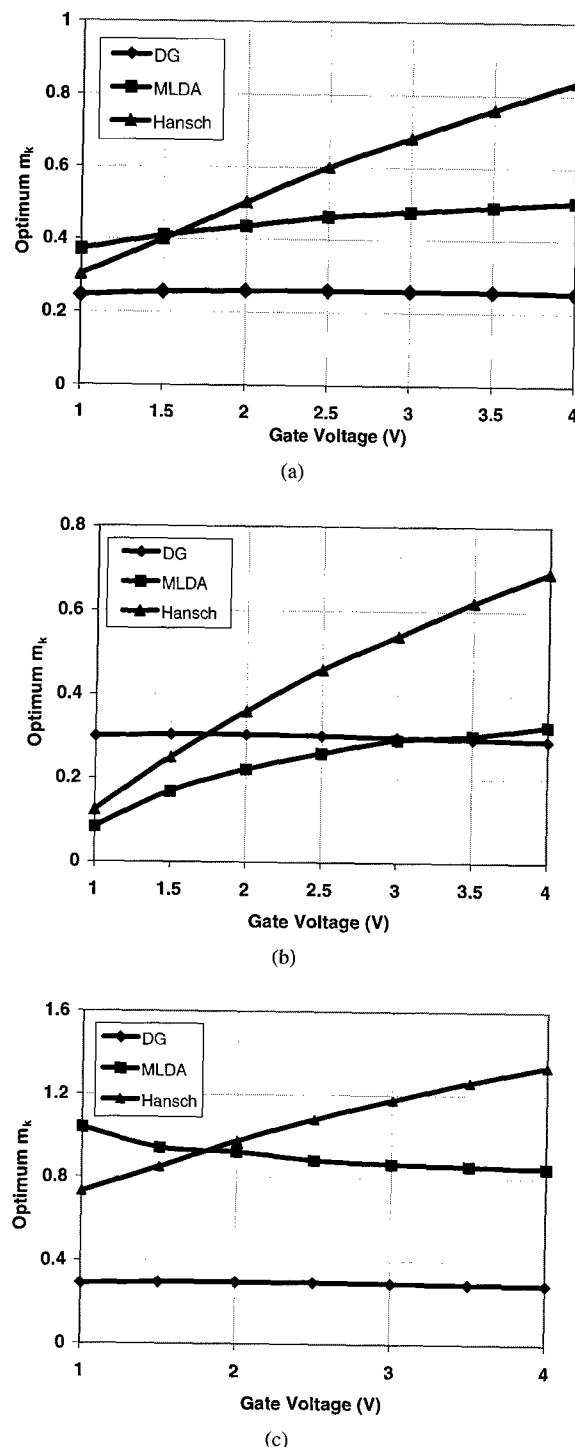


Figure 1. The optimum parameter m_k for the different structures based on different models. (a) Constant doping profile with $N_a = 1 \times 10^{17} \text{ cm}^{-3}$. (b) Low-high (retrograde) doping profile, with $N_s = 1 \times 10^{17} \text{ cm}^{-3}$ near interface and abruptly rising to $N_{bulk} = 1 \times 10^{18} \text{ cm}^{-3}$ at $x = 10 \text{ nm}$. (c) Gaussian doping profile with $N_{a0} = 2 \times 10^{17} \text{ cm}^{-3}$, $D_I = 2 \times 10^{12} \text{ cm}^{-2}$, and $R_p = \Delta R_p = 10 \text{ nm}$.

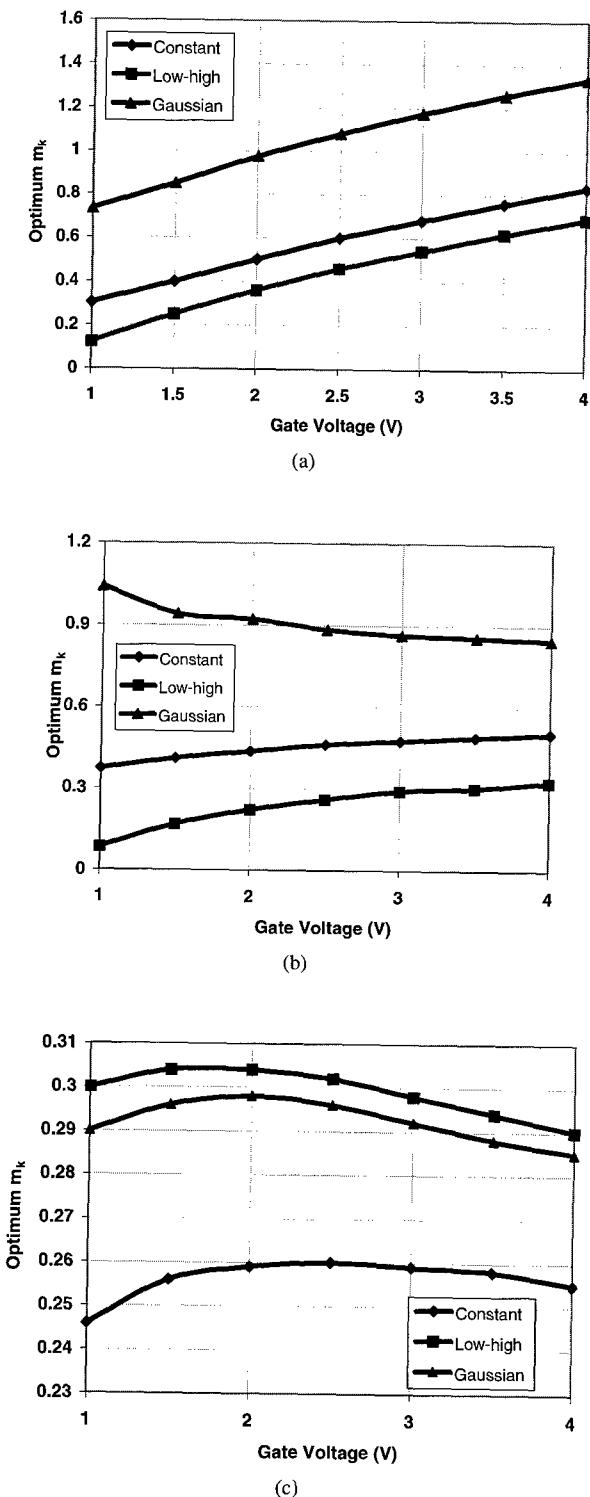


Figure 2. The optimum parameter m_k for the different structures based on different models. (a) Hansch model. (b) MLDA (modified local density approximation) model. (c) DG (density gradient) model.

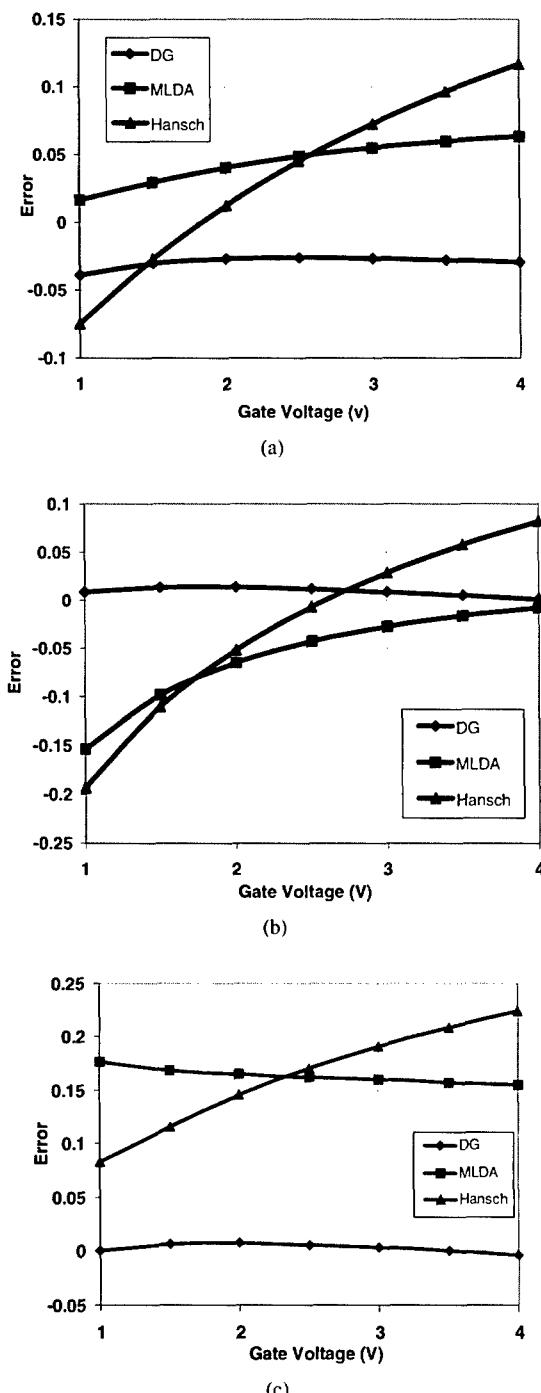


Figure 3. $\text{error} = |\bar{x} - \bar{x}_{s-p}|/\bar{x}_{s-p}$ vs. gate voltage by using the fixed optimum m_k obtained at $V_g = 2$ V and $N_A = 5 \times 10^{17} \text{ cm}^{-3}$ for different doping distributions in Si layer. (a) Constant doping profile with $N_a = 1 \times 10^{17} \text{ cm}^{-3}$. (b) Low-high (retrograde) doping profile, with $N_s = 1 \times 10^{17} \text{ cm}^{-3}$ near interface and abruptly rising to $N_{\text{bulk}} = 1 \times 10^{18} \text{ cm}^{-3}$ at $x = 10 \text{ nm}$. (c) Gaussian doping profile with $N_{a0} = 2 \times 10^{17} \text{ cm}^{-3}$, $D_I = 2 \times 10^{12} \text{ cm}^{-2}$, and $R_p = \Delta R_p = 10 \text{ nm}$.

with a standard ion-implantation process given by

$$N(x) = N_{a0} + \frac{D_I}{\sqrt{2\pi \Delta R_p}} \exp\left[-\frac{(x - R_p)^2}{2\Delta R_p^2}\right], \quad (7)$$

in which $N_{a0} = 2 \times 10^{17} \text{ cm}^{-3}$, $D_I = 2 \times 10^{12} \text{ cm}^{-2}$, and $R_p = \Delta R_p = 10 \text{ nm}$.

Using m_k as an adjustable parameter to best fit the “exact” solution, for the non-tunneling boundary condition at the interface, we have found that m_k for the D-G model is least sensitive to the substrate doping concentrations and applied gate voltages, followed next by the MLDA model. The Hänsch model is the worst, as shown in Figs. 1(a–c) and 2(a–c). If we choose the optimum parameter m_k at a uniform doping profile with $N_a = 5 \times 10^{17} \text{ cm}^{-3}$, $V_g = 2$ V as a reference and apply the same m_k to other doping profiles, it can be seen that D-G method introduces the smallest error among the three models, as shown in Fig. 3(a–c). However, the Hänsch model is simple and easy to implement, which can be used to calculate the initial guess for the other models. In terms of numerical computation, the MLDA model involves an extra integration but does not pose any convergence problem. Although the D-G model produces the solution closest to that of Schrödinger’s equation, special care is needed for the discretization scheme in order to be compatible with the boundary conditions (Wang 2001, Tang, Wang and Li 2001).

4. Conclusions

The Hänsch model and the MLDA approximation give an explicit expression for n_{QM} which can predict the S-P solution by adjusting the parameter m_k . But this parameter m_k is sensitive to the substrate dopings and applied voltages, especially for the Hänsch model. The D-G method does not give an explicit expression for n_{QM} . Instead, an additional perturbation term, which is often referred as ‘quantum diffusion’, is introduced in the continuity equation. Since a higher-order PDE with a singular perturbation term is involved, a special numerical treatment is needed for discretization. Erroneous results may be caused from using the non-linear fitting scheme with incompatible boundary conditions (Tang, Wang and Li 2001). There may be still room for improvement in the numerical solution scheme for the solution of D-G equations in multi-dimensions.

Acknowledgment

This work was supported in part by the National Science Foundation under NSF Grant E9710463 and ECS-0120128. The second author (T.-w. Tang) also acknowledges a support from the National Science Council through the National Center of High-Performance Computing in Hsinchu, Taiwan.

References

- Ancona M.G. 1997. Density-gradient simulations of quantum effects in ultra-thin-oxide MOS structures. JTCAD 97–100.
- Ancona M.G. and Biegel B.A. 2000. Nonlinear discretization scheme for the density-gradient equations. In: Proc. SISPAD'00, p. 196.
- Ancona M.G. and Tiersten H.F. 1987. Macroscopic physics of the silicon inversion layer. *Physical Review B* 35(15): 7959–7965.
- Hänsch W., Vogelsang T., Kircher R., and Orlowski M. 1989. Carrier transport near the Si/SiO₂ interface of a MOSFET. *Solid-State Electronics* 32(10): 839–849.
- Ieong M., Logan R., and Slinkman J. 1998. Efficient quantum correction model for multi-dimensional CMOS simulations. In: SISPAD'98, pp. 129–132.
- Paasch G. and Ubensee H. 1982. A modified local density approximation. *Phys. Stat. Sol. (b)* 113: 165–178.
- Tang T.-w., Wang X., and Li Y. 2001. Discretization scheme for the density-gradient equation and effect of boundary condition. In: IWCE-8.
- Wang X. 2001. Quantum correction to the charge density distribution in inversion layers. Master's Thesis, University of Massachusetts, Amherst.



Can the Density Gradient Approach Describe the Source-Drain Tunnelling in Decanano Double-Gate MOSFETs?

J.R. WATLING, A.R. BROWN AND A. ASENOV

*Device Modelling Group, Department of Electronics and Electrical Engineering, University of Glasgow,
Glasgow G12 8LT, Scotland, UK*

J.Watling@elec.gla.ac.uk

Abstract. As MOSFETs are scaled into the deep sub-micron (decanano) regime, quantum mechanical confinement and tunnelling start to dramatically affect their characteristics. It has already been demonstrated that the density gradient approach can be successfully calibrated in respect of vertical quantum confinement at the Si/SiO₂ interface and can reproduce accurately the quantum mechanical threshold voltage shift. In this paper we investigate the extent to which the density gradient approach can reproduce direct source-drain tunnelling in short double gate MOSFET devices.

Keywords: simulation, density gradient, tunnelling, double gate MOSFET

1. Introduction

As MOSFETs are scaled into the deep sub-micron regime, quantum mechanical (QM) confinement and tunnelling start to dramatically affect their characteristics. However, at present, complete quantum simulations involving, for example, Wigner or Green's functions are expensive and therefore not suitable for inclusion within CAD simulation tools. The common practice is therefore to introduce economical first-order quantum corrections into conventional drift-diffusion simulators. This can be accomplished using the well-established density gradient (DG) formalism (Ancona and Iafrate 1989).

In this paper we use a double gate MOSFET with simple architecture and gate lengths in the range 30 to 6 nm as vehicle for this study. The paper investigates the extent to which the DG approach can reproduce the phenomena of source-drain (S-D) tunnelling in extremely short devices. Experimental evidence for S-D tunnelling has been observed (Kawaura *et al.* 2000) and manifests itself as a degradation of the subthreshold current slope and anomalous temperature dependence. Properly scaled conventional MOSFETs with 20 nm channel lengths have already been demonstrated by leading semiconductor

manufacturers (Chau 2001). It is, however, common wisdom that the scaling of the field effect transistor below this milestone requires intolerably thin gate oxides and unacceptably high channel doping, therefore advocating a departure from the conventional MOSFET concept. One of the most promising new device structures, scalable to dimensions of 10 nm and below is the double or wrap around gate MOSFET. Thus it is likely that in these structures direct S-D tunnelling may become significant. Additionally in a double gate structure the current is essentially one-dimensional, making theoretical study and calibration easier than in a conventional MOSFET device structure.

The next section describes the double gate MOSFET structure considered in this work. Section 3, describes the now well-established DG formalism, and to what extent this approach may include tunnelling. Our results and evidence for source-drain tunnelling are presented in Section 4, while Section 5 presents our conclusion and discussions.

2. Double-Gate Structure

Here we have studied an archetypal double gate MOSFET structure, similar in design to that by Ren *et al.* (2000). We have investigated a family of

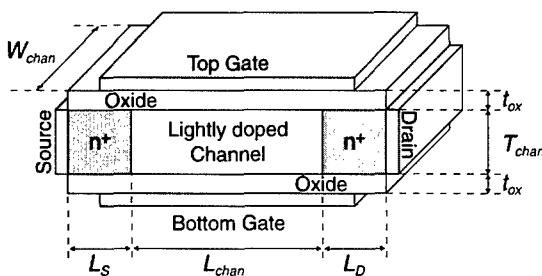


Figure 1. Schematic representation of the double-gate MOSFET structure considered in this work.

double-gate MOSFETs illustrated schematically in Fig. 1, with channel lengths, L_{chan} , ranging from 30 nm down to 6 nm, with a width, W_{chan} , of 30 nm. The channel thickness, T_{chan} , and oxide thickness, t_{ox} , are both fixed at 1.5 nm. The source and drain junctions are 10 nm and doped at $1 \times 10^{20} \text{ cm}^{-3}$ the channel is lightly doped at $1 \times 10^{16} \text{ cm}^{-3}$.

It is this lightly doped channel, that makes the double gate structure resilient to random dopant fluctuations, which we have studied in another paper (Brown, Watling and Asenov to be published).

3. Density-Gradient Formalism

The density gradient method may be derived from the one particle Wigner function (Carruthers and Zachariasen 1983):

$$\frac{\partial f(\mathbf{k}, \mathbf{r}, t)}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f(\mathbf{k}, \mathbf{r}, t) - \frac{2}{\hbar} V(\mathbf{r}) \sin \left[\frac{\hbar \vec{\nabla}_{\mathbf{r}} \vec{\nabla}_{\mathbf{k}}}{2} \right] \\ \times f(\mathbf{k}, \mathbf{r}, t) = \left(\frac{\partial f(\mathbf{k}, \mathbf{r}, t)}{\partial t} \right)_{coll} \quad (1)$$

Quantum effects are included through the inherently non-local driving potential in the third term on the left-hand side. Expanding to first order in \hbar , so that only the first non-local quantum term is considered, has been shown to be sufficiently accurate to model non-equilibrium quantum transport and also for the inclusion of tunnelling phenomena in particle based Monte Carlo simulators (Tsuchiya and Miyoshi 2000, Tsuchiya, Fischer and Hess 2000). The additional, non-classical, quantum correction term may be viewed as a modification to the classical potential and acts like an additional quantum force term in the particle simulations, similar in spirit to the Bohm interpretation. In the

case of a tunnelling barrier, the additional term acts to raise the classical conduction band potential profile to the left of the barrier and lower the classical potential barrier, for carriers flowing from left to right.

The density gradient approximation maybe derived in a manner similar to that for deriving the drift diffusion approximation for the Boltzmann Transport Equation (Snowden 1989). The classical electronic equation of state is thus modified so that it includes an additional term that is dependent on the gradient of the carrier density:

$$2b_n \frac{\nabla^2 \sqrt{n}}{\sqrt{n}} = \phi_n - \psi + \frac{k_B T}{q} \ln \left(\frac{n}{n_i} \right) \quad (2)$$

where

$$b_n = \frac{\hbar^2}{12 q m_n^*}.$$

It remains unclear however, if the approximations required in deriving the DG approach remove the ability to be able to model tunnelling phenomena, there stills remains controversy over whether the DG and other similar approaches such as effective potential (Ferry, Akis and Vasileska 2000) can model tunnelling. However, it is clear that the DG formalism will be unable to cope with cases where tunnelling is dependent on the coherent phase behaviour of electrons, as in the case of resonant tunnelling. We may therefore consider that DG, if it can account for tunnelling, does so in what may be termed the scattering-dominated limit (Ancona 2001). Here, we have perform a series of numerical experiments to see if DG can, at least qualitatively, account for the impact of source-drain tunnelling on the subthreshold $I_D - V_G$ characteristics of very short double gate MOSFETs.

4. Results

It has already been demonstrated (Asenov *et al.* 2001, Watling *et al.* 2001) that the DG approach can be successfully calibrated in respect of vertical quantum confinement at the Si/SiO₂ interface and can reproduce accurately the QM threshold voltage shift by adjusting the effective mass in the vertical direction. An effective mass of $0.19m_0$ is found to give the best agreement. Here we investigate through a variety of numerical experiments the extent to which the DG approach reproduces at least qualitatively the impact of source-drain

tunnelling on the I_D - V_G characteristics and the subthreshold slope in short devices and possibly calibrated by means of the lateral effective mass.

The short channel lengths and channel thickness means that quantum effects become significant, thus making the use of classical simulations untrustworthy. It is therefore mandatory to include quantum corrections, such as through the DG formalism. The significance of the quantum effects can clearly be seen in Figs. 2 and 3.

Figures 2 and 3, show the corresponding classical and quantum charge density profiles respectively, in the direction normal to the gate, it can be seen that the quantum distribution tends to zero at the Si/SiO₂ interface while the classical distribution peaks at the Si/SiO₂ interface.

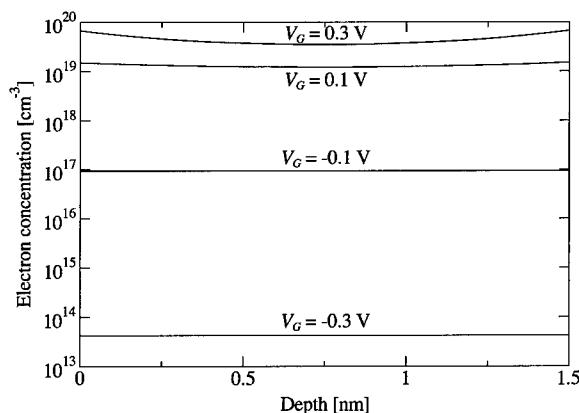


Figure 2. Classical electron concentration profile through the centre of a $30 \times 30 \times 1.5$ nm double-gate MOSFET.

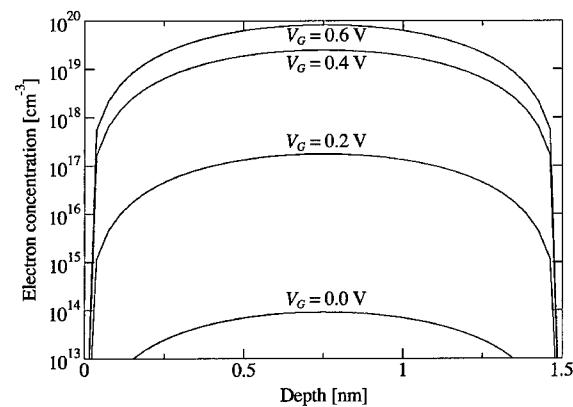


Figure 3. Quantum (density gradient) electron concentration profile through the centre of a $30 \times 30 \times 1.5$ nm double-gate MOSFET.

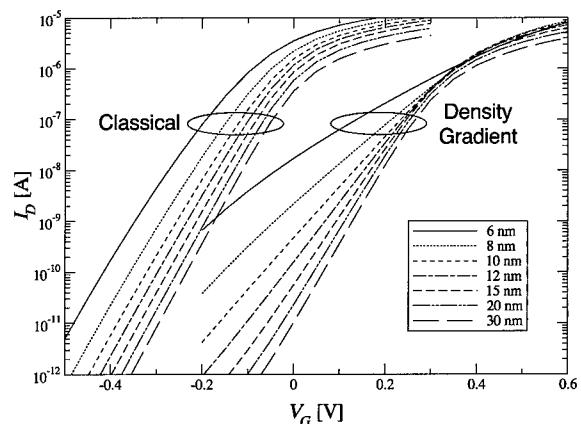


Figure 4. I_D - V_G characteristics for a double gate structure, with gate lengths ranging from 30 nm down to 6 nm, obtained from our classical and density gradient simulations. $V_D = 0.01$ V and V_G is applied to both top and bottom gate contacts.

The quantum confinement effects in the DG case leads to a large quantum mechanically threshold voltage shift, (~ 0.3 V for the 10 nm channel device), shown in Fig. 4.

Using a constant value ($0.19m_0$) for the effective mass in all directions, we observe that the subthreshold slope in the DG simulations degrades significantly as the channel length is decreased, while in the classical simulations the subthreshold slope remains nearly constant with channel length. For a channel length of 30 nm the classical and DG subthreshold slopes are almost identical. However, as the gate length is shrunk down to 10 nm and below the subthreshold slope degradation in the DG simulations becomes significant as has been observed by other researchers (Lundstrom 2001). All of these observations provide an indication that S-D tunnelling, is included to some extent in the DG simulations.

Further evidence can be gain by looking at the temperature dependence of the subthreshold slope. Classical MOSFET theory dictates that the classical subthreshold slope S is given by Taur and Ning (1998):

$$S = \left(\frac{d(\log_{10} I_{ds})}{dV_g} \right)^{-1} \approx 2.3 \frac{k_B T}{q} \left(1 + \frac{C_{dm}}{C_{ox}} \right) \quad (3)$$

Thus the classical subthreshold slope depends linearly on temperature, as we would expect as the classical subthreshold current is essentially thermionic in nature, so it has an approximately exponential dependence. However, any current due to tunnelling will have

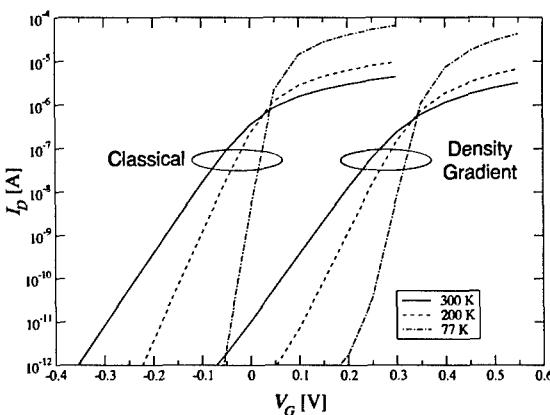


Figure 5. I_D - V_G characteristics for a 30 nm channel length double gate structure from classical and density gradient simulations, for a range of temperatures. $V_D = 0.01$ V and V_G is applied to both top and bottom gate contacts.

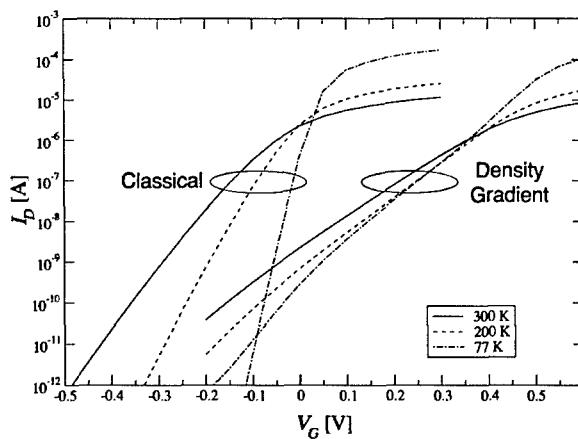


Figure 6. I_D - V_G characteristics for an 8 nm channel length double gate structure from classical and density gradient simulations, for a range of temperatures. $V_D = 0.01$ V and V_G is applied to both top and bottom gate contacts.

a much weaker dependence on temperature (Kawaura *et al.* 2000). Figures 5 and 6, show the temperature dependence of the subthreshold slope in both classical and DG simulations, for channel lengths of 30 nm and 8 nm respectively.

We observe here that the temperature dependence of the subthreshold slope is similar for both the classical and DG simulations in the 30 nm gate length device, in agreement with Eq. (3). The shift in the I_D - V_G is the QM threshold voltage shift caused by quantization in the vertical direction, as illustrated in Figs. 3 and 4. However, for an 8 nm gate length device, there is a

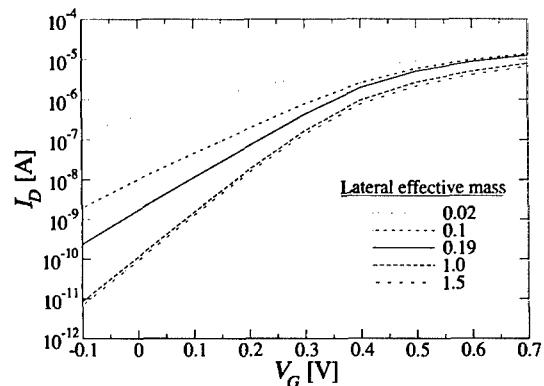


Figure 7. I_D - V_G characteristics for an 8 nm channel length double gate structure obtained from our density gradient simulations, at 300 K for different lateral effective mass. $V_D = 0.01$ V and V_G is applied to both top and bottom gate contacts.

noticeable degradation of the subthreshold slope in the DG simulations as compared with the classical simulations, indicating the existence of a second current transport mechanism in subthreshold region in addition to the classical over-barrier (thermionic) current. This is further supported by the observation that in the DG simulations the subthreshold slope is nearly independent of temperature. These observations are again consistent with the possibility of a source-drain tunnelling current, which is less sensitive to temperature than a thermionic emission current.

All the results presented so far have been of a qualitative nature, as we have assumed the same effective mass in the lateral direction, as in the vertical direction. However, the lateral effective mass would need to be calibrated in respect of source-drain tunnelling, in order to be able to perform quantitative simulations. We have therefore performed simulations, where we have varied the lateral effective mass, shown in Fig. 7.

We observe that the increase of the lateral effective mass results in an increase in the subthreshold slope as expected since the equation of state in the DG formalism (Eq. (2)) becomes more classical-like in the lateral direction. There is also a slight shift in the threshold voltage caused by a mixing effect of the effective mass in the vertical and lateral directions.

5. Conclusion

We have performed a variety of numerical experiments to investigate whether the density gradient approach

can model source-drain tunnelling in double gate MOSFETs in respect of the subthreshold current characteristics in decanano scale MOSFETs. A variety of double gate MOSFETs, with channel lengths ranging from 30 nm to 6 nm have been studied. We observe that as the channel length is reduced, there is a corresponding reduction in the subthreshold slope, in line with the available experimental evidence (Kawaura *et al.* 2000). The temperature dependence of the subthreshold slope has also been studied, it is observed that temperature dependence of the 30 nm MOSFET is in agreement with standard MOSFET theory, while the subthreshold slope for the 8 nm device is nearly independent of temperature, presumably due to the larger source-drain tunnelling in the smaller device, which is less temperature sensitive than the classical thermionically dominated subthreshold current. All of these facts are in agreement with experimental observations of direct-source drain tunnelling.

While it remains an open question whether density gradient can describe quantitatively the tunnelling phenomena, the series of computational experiments performed here provide evidence that, at least qualitatively, this approach can reproduce the important aspects of the I_D - V_G characteristics that are consistent with the presence of source-drain tunnelling.

Calibration of the vertical and lateral effective mass, with respect to both the quantum mechanical threshold voltage shift and source-drain tunnelling, may make it possible to perform quantitative quantum simulations, using density gradient. However, this may be difficult as there is clearly some mixing between the lateral and vertical masses, as revealed in our simulations.

Acknowledgments

We gratefully acknowledge the helpful and useful discussions with Prof. Mark Lundstrom, Dr. Dejan Jovanovic, Dr. Mario Ancona and Dr. Anant Anantram. JRW would like to acknowledge the support of EPSRC under grant no GR/L53755. SHEFC Research Development Grant VIDEOS provided support for ARB.

References

- Ancona M. 2001. Private communication.
- Ancona M.G. and Iafrate G.J. 1989. Physical Review B 39: 9536–9540.
- Asenov A., Slavcheva G., Brown A.R., Davies J.H., and Saini S. 2001. IEEE Trans. Electron Devices 48: 722–729.
- Brown A.R., Watling J.R., and Asenov A. Proceedings of IWCE-8, to be published.
- Carruthers P. and Zachariasen F. 1983. Review of Modern Physics 55: 245–284.
- Chau R. 2001. Si Nanoelectronics Workshop, pp. 2–3.
- Ferry D.K., Akis R., and Vasileska D. 2000. IEDM Tech. Digest, pp. 287–290.
- Kawaura H., Sakamoto T., Baba T., Ochiai Y., Fujita J., and Sone J. 2000. IEEE Trans. Electron Devices 47: 856–860.
- Lundstrom M. 2001. Private communication.
- Ren Z., Venugopal R., Datta S., and Lundstrom M. 2000. IEDM Technical Digest, pp. 715–718.
- Snowden C.M. 1989. Semiconductor Device Modelling. Springer-Verlag, Wien, New York.
- Taur Y. and Ning T.H. 1998. Fundamentals of MODERN VLSI DEVICES. Cambridge University Press.
- Tsuchiya H., Fischer B., and Hess K. 2000. IEDM Tech. Digest, pp. 283–286.
- Tsuchiya H. and Miyoshi T. 2000. Superlattices and Microstructures 27: 529–532.
- Watling J.R., Brown A.R., Asenov A., and Ferry D.K. 2001. In: Proc. SISPAD, pp. 82–85.



A Particle Description Model for Quantum Tunneling Effects

HIDEAKI TSUCHIYA*

Department of Electrical and Electronics Engineering, Kobe University, Japan

tsuchiya@eedept.kobe-u.ac.jp

UMBERTO RAVAIOLI

*Beckman Institute and Department of Electrical and Computer Engineering, University of Illinois
at Urbana-Champaign, USA*

Abstract. We present here a particle description model for quantum tunneling effects. A quantum force has been formulated based on a truncation to first order of the expansion form of the Wigner transport equation, and has been incorporated into the semiclassical Monte Carlo simulation. The combined Monte Carlo/quantum force approach was applied to simulations for resonant-tunneling effects.

Keywords: quantum force correction, Monte Carlo simulation, Wigner transport equation, tunneling

1. Introduction

In usual quantum approaches, the physical state of an individual system is specified by a wave function obtained from the solution of Schrödinger equation. For practical device simulations at normal temperatures, the use of a full quantum wave theory is still problematic because of the difficulty in including realistic scattering models. In alternative, a particle description of quantum theory is possible, in terms of a quantum potential/force correction (Bohm 1952, Tsuchiya and Ravaioli 2001). In this case, the notion of a well-defined particle trajectory is retained, while the quantum force correction modifies the potential energy profile to account for quantum effects. A particle-based approach coupled with quantum force correction is very attractive for practical simulation of nanoscale semiconductor devices (Tsuchiya and Ravaioli 2001). We present here a quantum correction approach derived from a simplification of the Wigner transport equation where the dynamics of particles can be treated as in semiclassical Monte Carlo (MC) simulation with a nonlocal quantum force. The model is

applied to MC particle simulation of resonant-tunneling effects.

2. Quantum-Corrected Monte Carlo

The transport equation for the Wigner distribution function is given in the form of a modified Boltzmann transport equation (BTE) as Wigner (1932)

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f - \frac{1}{\hbar} \nabla_{\mathbf{r}} U \cdot \nabla_{\mathbf{k}} f + \sum_{\alpha=1}^{\infty} \frac{(-1)^{\alpha+1}}{\hbar^{4\alpha} (2\alpha+1)!} (\nabla_{\mathbf{r}} \cdot \nabla_{\mathbf{k}})^{2\alpha+1} U f = \left(\frac{\partial f}{\partial t} \right)_C \quad (1)$$

where U denotes the spatially varying potential energy. Note that $\nabla_{\mathbf{k}}$ operates only on f and $\nabla_{\mathbf{r}}$ operates only on the potential U . An essence of the Wigner formalism is the presence of quantum corrections through the inherently nonlocal driving potential, in the expansion of the fourth term on the left-hand side of Eq. (1). Here, we indicate with Q_1 the lowest-order quantum correction term obtained by considering only $\alpha = 1$ in the expansion of Eq. (1). The lowest-order term gives a major contribution in the quantum mechanical corrections.

*To whom correspondence should be addressed.

For a three-dimensional problem, Q_1 is written, as

$$Q_1 = \frac{1}{24\hbar} \left(\frac{\partial^3 U}{\partial x^3} \frac{\partial^3 f}{\partial k_x^3} + \frac{\partial^3 U}{\partial y^3} \frac{\partial^3 f}{\partial k_y^3} + \frac{\partial^3 U}{\partial z^3} \frac{\partial^3 f}{\partial k_z^3} + 3 \frac{\partial^3 U}{\partial x^2 \partial y} \frac{\partial^3 f}{\partial k_x^2 \partial k_y} + 3 \frac{\partial^3 U}{\partial x^2 \partial z} \frac{\partial^3 f}{\partial k_x^2 \partial k_z} + \dots \right) \quad (2)$$

Supposing that the system is relatively close to equilibrium, we introduce for simplicity a displaced Maxwell-Boltzmann distribution in Eq. (2), as $f = \exp\{-\beta[E_{\mathbf{k}-\bar{\mathbf{k}}} + U(\mathbf{r}) - E_f]\}$, where E_f is the Fermi energy, $\beta = 1/k_B T$, $E_{\mathbf{k}-\bar{\mathbf{k}}}$ the carrier's energy and $\bar{\mathbf{k}}$ the average momentum of the displaced distribution function. By using the above distribution function with the effective mass approximated carrier's energy, we can obtain relations, as Tsuchiya and Ravaioli (2001)

$$\frac{\partial f}{\partial k_x} = -\frac{\beta\hbar^2}{m_x}(k_x - \bar{k}_x)f = -\gamma_x(k_x - \bar{k}_x)f \quad (3)$$

$$\frac{\partial^2 f}{\partial k_x^2} = \gamma_x f [\gamma_x(k_x - \bar{k}_x)^2 - 1] \quad (4)$$

$$\frac{\partial^3 f}{\partial k_x^3} = [\gamma_x^2(k_x - \bar{k}_x)^2 - 3\gamma_x] \frac{\partial f}{\partial k_x} \quad (5)$$

and

$$\frac{\partial^3 U}{\partial x^3} = -\frac{1}{\beta} \frac{\partial^3 \ln(n)}{\partial x^3} \quad (6)$$

where n is the carrier density and $\gamma_x = \beta\hbar^2/m_x$. Equations (4) and (5) are obtained by using Eq. (3). Similarly, we can express the other terms of Eq. (2), and then obtain a quantum-corrected BTE, as Tsuchiya and Ravaioli (2001)

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \frac{1}{\hbar} (-\nabla_{\mathbf{r}} U + \mathbf{F}^Q) \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t} \right)_C \quad (7)$$

Quantum effects are incorporated in terms of quantum mechanical driving forces $\mathbf{F}^Q = (F_x^Q, F_y^Q, F_z^Q)$, as

$$F_x^Q = \frac{\partial}{\partial x} \left\{ \frac{-1}{24\beta} \left[(\gamma_x^2 \Delta k_x^2 - 3\gamma_x) \frac{\partial^2}{\partial x^2} + 3(\gamma_y^2 \Delta k_y^2 - \gamma_y) \frac{\partial^2}{\partial y^2} + 3(\gamma_z^2 \Delta k_z^2 - \gamma_z) \frac{\partial^2}{\partial z^2} \right] \ln(n) \right\} \quad (8)$$

$$F_y^Q = \frac{\partial}{\partial y} \left\{ \frac{-1}{24\beta} \left[(\gamma_y^2 \Delta k_y^2 - 3\gamma_y) \frac{\partial^2}{\partial y^2} + 3(\gamma_x^2 \Delta k_x^2 - \gamma_x) \frac{\partial^2}{\partial x^2} + 3(\gamma_z^2 \Delta k_z^2 - \gamma_z) \frac{\partial^2}{\partial z^2} \right] \ln(n) \right\} \quad (9)$$

$$F_z^Q = \frac{\partial}{\partial z} \left\{ \frac{-1}{24\beta} \left[(\gamma_z^2 \Delta k_z^2 - 3\gamma_z) \frac{\partial^2}{\partial z^2} + 3(\gamma_x^2 \Delta k_x^2 - \gamma_x) \frac{\partial^2}{\partial x^2} + 3(\gamma_y^2 \Delta k_y^2 - \gamma_y) \frac{\partial^2}{\partial y^2} \right] \ln(n) \right\} \quad (10)$$

where $\gamma_i = \beta\hbar^2/m_i$ and $\Delta k_i = k_i - \bar{k}_i$ ($i = x, y, z$). \bar{k}_i is again the average momentum of the distribution function depending on the position.

The momentum components, k_x , k_y , and k_z are explicitly included in Eqs. (8)–(10). An additional approximation can be made by assuming a thermal equilibrium energy for the momentum terms as $\hbar^2(k_i - \bar{k}_i)^2/2m_i \approx k_B T/2 = 1/2\beta$. Then, we obtain the relation of $\Delta k_i^2 \approx 1/\gamma_i$, and the corresponding quantum forces are simply represented by

$$F_x^Q = \frac{\partial}{\partial x} \left(\frac{\hbar^2}{12m_x} \frac{\partial^2 \ln(n)}{\partial x^2} \right) \quad (11)$$

$$F_y^Q = \frac{\partial}{\partial y} \left(\frac{\hbar^2}{12m_y} \frac{\partial^2 \ln(n)}{\partial y^2} \right) \quad (12)$$

$$F_z^Q = \frac{\partial}{\partial z} \left(\frac{\hbar^2}{12m_z} \frac{\partial^2 \ln(n)}{\partial z^2} \right) \quad (13)$$

This formulation differs from the results in Eqs. (8)–(10) in the fact that it gives a force which depends only on the position but not on the momentum of particles. The simplified quantum forces given by Eqs. (11)–(13) could be useful, for instance, for simulation of size-quantization in the channel of ultra-small MOSFETs.

Based upon Eq. (7), the velocity and the force for particles during free flights are given, respectively, as

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{k}}{dt} = \frac{1}{\hbar} (-\nabla_{\mathbf{r}} U + \mathbf{F}^Q) \quad (14)$$

The velocity equation is the same as used in the standard MC technique, but the force equation is modified so that the particles evolve under the influence of the classical driving force $-\nabla_{\mathbf{r}} U$, plus the quantum force \mathbf{F}^Q . Qualitatively, the effect of the quantum force correction is to

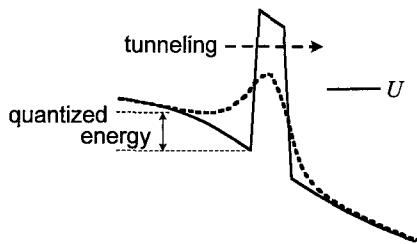


Figure 1. Qualitative effects of quantum force correction for single barrier structure. The solid line and the dashed line denote the classical potential and quantum-corrected potential, respectively.

smooth out sharp changes in the potential as shown in Fig. 1, where the solid line and the dashed line denote the classical potential and quantum-corrected potential, respectively. Consequently, the tunneling and quantum confinement effects can be incorporated in semiclassical carrier transport models. Note that we do not solve the Schrödinger equation and do not introduce a wave packet representation in quantum structures (Baba, Al-Mudares and Barker 1989, Oriols *et al.* 1998). A full particle description of quantum processes could be attempted in practical simulations.

3. Resonant-Tunneling Simulation

We present here the results of computational experiments for resonant-tunneling particles. We consider a double-barrier structure consisting of GaAs and AlGaAs, where quantum interference effects can be carefully identified. In the calculations, we used

$$F_x^Q = \frac{\partial}{\partial x} \left\{ \frac{-1}{24\beta} [\gamma_x^2 (k_x - \bar{k}_x)^2 - 3\gamma_x] \frac{\partial^2 \ln(n)}{\partial x^2} \right\} \quad (15)$$

which corresponds to a one-dimensional version of Eq. (8)–(10). The barrier height and width are 0.22 eV and 2.5 nm, respectively, and the quantum well width is 4.5 nm. We simulate the electron transport in the Γ valley at room temperature (300 K). The doping density in the GaAs electrodes is taken to be 10^{18} cm^{-3} . As scattering processes we consider LO phonons, acoustic phonons, and ionized impurity scatterings. Figure 2 shows a snapshot of the computed electron distributions in space and energy at zero bias voltage, where (a) corresponds to the classical MC simulation and (b) to the quantum-corrected MC simulation. For reference, the conduction band profiles are also plotted with solid lines. Note that the vertical axis denotes the total

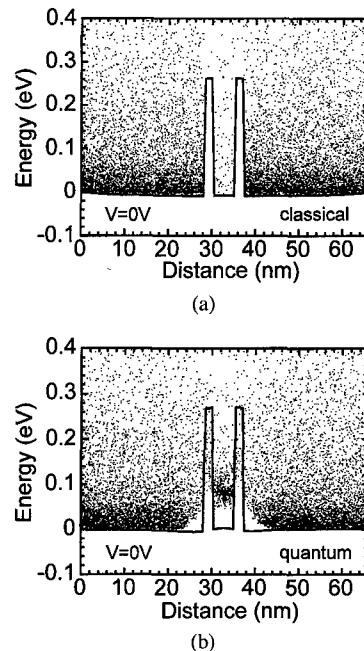


Figure 2. Electron distributions in space and energy of double-barrier resonant-tunneling structure at zero bias voltage. (a) Corresponds to the classical MC simulation without quantum force and (b) to the quantum-corrected MC simulation with quantum force. The conduction band profiles are also plotted with solid lines, and the vertical axis denotes the total electron energy including the contribution of quantum force (quantum potential).

electron energy, including the contribution of quantum force (15). In Fig. 2(b), the quantum tunneling particles are found inside the potential barriers, in addition to the thermally excited ones. This is because the potential barrier is effectively lowered due to the quantum force correction as explained in Fig. 1. We can also observe the formation of quantized subbands in the central quantum well. The quantum force correction prevents the electrons from occupying energy states below a certain level, as imposed by the formation of quantized subbands in the well. We estimated from the particle distribution in Fig. 2(b) the electron's energy distribution function confined in the quantum well to compare with a tunneling probability. Figure 3 shows the estimated distribution function of electrons in the quantum well, corresponding to Fig. 2(b). The dashed line indicates the corresponding tunneling probability as a function of energy calculated by using a transfer matrix solver of the Schrödinger equation. For the transfer matrix calculation, we used the potential distribution data obtained from the MC simulation. The peak

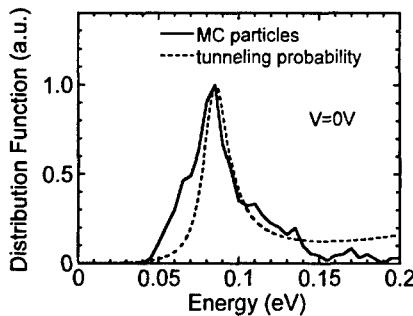


Figure 3. Distribution function of electrons in quantum well. The dashed line indicates corresponding tunneling probability calculated by using a transfer matrix solver of Schrödinger equation.

energy corresponds to the quantized energy level in the well and the shape of the function denotes the resonant energy broadening. We can see that both methods predict the identical quantized energy of 85 meV. For the resonant energy broadening, the MC simulation gives slightly broader result, especially in the lower energy region. This should be due to the phonon emission scattering of electrons, which is effectively included in the quantum-corrected MC results. Since the system is at the thermal equilibrium for Figs. 2 and 3, the quantum-corrected MC and transfer matrix results are in good agreement, although the MC results fluctuate somewhat in the higher energy region due to the discreteness of the particle energy distribution.

Next, we present a nonequilibrium simulation with applied external bias. Figure 4 shows the computed electron distribution in space and energy at a bias voltage of 0.13 V. The particle distribution confined in the quantum well is found to shift toward the lower energy side than that in Fig. 2(b), due to the influence of bias

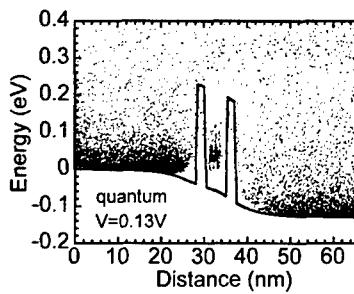
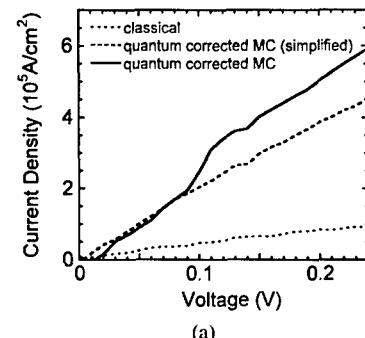
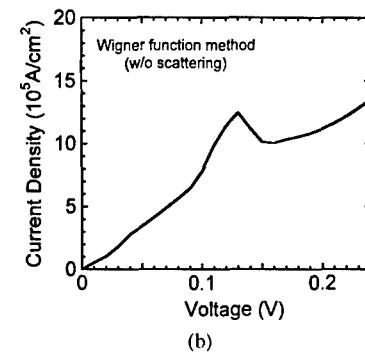


Figure 4. Electron distribution in space and energy at 0.13 V simulated by using the quantum-corrected MC method. The conduction band profile is also plotted with solid line, and the vertical axis denotes the total electron energy including the contribution of quantum force (quantum potential).



(a)



(b)

Figure 5. Simulated current-voltage characteristics. (a) Corresponds to the MC simulation results and (b) to the full Wigner function method without scattering. The solid line and the dotted line in (a) correspond to the quantum-corrected MC and the classical MC results, respectively. For comparison, we plotted with the dashed line the result calculated by using the simplified quantum force model (11).

voltage. This corresponds to the downward shift of tunneling probability spectrum when the bias is applied. Since the current flows largely in the case of Fig. 4, the confined particle concentration becomes fewer than that at thermal equilibrium (Fig. 2(b)).

Figure 5 shows the simulated current-voltage characteristics, where (a) indicates the results from the MC simulations and (b) from the full Wigner function method without scattering. The solid line and the dotted line in Fig. 5(a) correspond to the quantum-corrected MC and the classical MC results, respectively. For comparison, we plotted with the dashed line the result calculated by using the simplified quantum force model (11). The quantum-corrected MC model (the solid line) indicates a step-like nonlinear behavior around $V = 0.13$ V. The similar behavior is weakly visible also in the simplified model (the dashed line). For the present device structure, the current peak is expected to appear at 0.13 V from the ballistic simulation result as shown in Fig. 5(b). The nonlinear curve in

Fig. 5(a) results from the double barrier structure. The quantum-corrected semiclassical simulation should be capturing the limit of sequential tunneling, in the presence of strong scattering. This points to need for more simulations at lower temperatures, where phonon scattering decreases, to characterize completely the behavior of the present model when transport transfers from sequential to resonant tunneling. Further model development may be needed in conjunction with careful comparisons with ballistic quantum results as in Fig. 5(b).

4. Conclusion

We have presented a particle description model for quantum tunneling effects based upon the Wigner's transport formalism, where the dynamics of particles can be treated as in semiclassical Monte Carlo simulation with a quantum force correction. The model was applied to the simulation of particle transport in double barrier structures where resonant tunneling is present. The simulations show the presence of confined

particles inside the quantum well and a nonlinear I - V characteristics at 300 K was obtained under strong scattering conditions. Further investigations at low temperatures will be required to understand how resonant-tunneling phenomena may be accounted for completely in particle simulations.

Acknowledgments

This work was supported by the Ministry of Education, Science, Sports and Culture of Japan, Grand-in-Aid for Encouragement of Young Scientists, 13750061, 2001.

References

- Baba T., Al-Mudares M., and Barker J.R. 1989. Japanese Journal of Applied Physics 28: L1682.
- Bohm D. 1952. Physical Review 85: 166.
- Oriols X., García-García J.J., Martín F., Suñé J., González T., Mateos J., and Pardo D. 1998. Applied Physics Letters 72: 806.
- Tsuchiya H. and Ravaioli U. 2001. Journal of Applied Physics 89: 4023.
- Wigner E. 1932. Physical Review 40: 749.

JOURNAL OF COMPUTATIONAL ELECTRONICS

Instructions for Authors

Authors are encouraged to submit high quality, original work that has neither appeared in, nor is under consideration by, other journals.

PROCESS FOR SUBMISSION

1. Authors should submit five hard copies of their final manuscript to:

Ms. Jesikah Allison
JOURNAL OF COMPUTATIONAL
ELECTRONICS
Editorial Office
Kluwer Academic Publishers
101 Philip Drive
Norwell, MA 02061
Tel.: 781-871-6600
FAX: 781-878-0449
Email: jesikah.allison@wkap.com

For prompt attention, all correspondence can be directed to this address.

2. Enclose with each manuscript, on a separate page, from five to ten index terms (key phrases).
3. Enclose originals for the illustrations (see STYLE FOR ILLUSTRATIONS section below). Alternatively, good quality copies may be sent initially, with the originals ready to be sent immediately upon acceptance of paper.
4. Enclose a separate page giving your complete preferred mailing address (including street and bldg. nos.) for correspondence and return of proofs. Please be sure to include a telephone number, fax number and e-mail address.
5. The refereeing is done by anonymous reviewers.
6. All papers should be written in English.

STYLE FOR MANUSCRIPT

1. Typeset, double or 1 1/2 space; use one side of sheet only (laser printed, typewritten and good quality duplication acceptable).

2. Provide an informative 100 to 250-word abstract at the head of the manuscript. The abstracts are printed with the articles.
3. Provide a separate double-spaced sheet listing all footnotes, beginning with "Affiliation of author" and continuing with numbered references. Acknowledgement of financial support may be given if appropriate.
4. References should appear in a separate bibliography at the end of the paper in alphabetical order with items referred to in the text by author and date of publication in parentheses, e.g. (Marr 1982). References should be complete, in the following style:

Style for papers: Authors, last names followed by first initials, year of publication, title, volume, inclusive page numbers.

Style for books: Authors, year of publication, title, publisher and location, chapter and page numbers (if desired).

Examples as follows:

(Book) Marr D. 1992, Vision, A Computational Investigation into the Human Representation & Processing of Visual Information, San Francisco, Freeman.

(Journal Article) Rosenfeld A., and Thurston M. 1971. Edge and curve detection for visual scene analysis, IEEE Trans. Comput. C-20: 562-569.

(Conference Proceedings) Witkin A. 1983. Scales space filtering. Proc. Int. Joint Conf. Artif. Intell., Karlsruhe, West Germany, pp. 1019-1021.

(Lab. memo.) Yuille A.L. and Poggio T. 1983. Scaling theorems for zero crossings. M.I.T. Artif. Intell. Lab., Mass. Inst. Tech., Cambridge, MA, A.I. Memo. 722.

5. Provide a separate sheet listing all figure captions, in proper style for the typesetter, e.g., "Fig. 3. Examples of the zero crossings of the second derivative

of the (a) Gaussian and (b) sine filter for the same input function."

ELECTRONIC DELIVERY

Please send only the electronic version (of ACCEPTED paper) via one of the methods listed below. Note, in the event of minor discrepancies between the electronic version and hard copy, the electronic file will be used as the final version.

Via electronic mail

1. Please e-mail electronic version to:
KAPfiles@wkap.com
2. Recommended formats for sending files via e-mail:
 - a. Binary files - uuencode or binhex
 - b. Compressing files - compress, pkzip or gzip
 - c. Collecting files - tar
3. The e-mail message should include the author's last name, the name of the journal to which the paper has been accepted, and the type of file (e.g., LaTeX or ASCII).

Via anonymous FTP

ftp: *ftp.wkap.com*

cd: /incoming/production

Send e-mail to *KAPfiles@wkap.com* to inform Kluwer electronic version is at this FTP site.

Via disk

1. Label a 3.5 inch floppy disk with the operating system and word processing along with the authors' names, manuscript title, and name of journal to which the paper has been accepted.
2. Mail disk to:

Kluwer Academic Publishers
Desktop Department
101 Philip Drive, Assinippi Park
Norwell, MA 02061, USA

Any questions about the above procedures please send e-mail to: *dthelp@wkap.com*

STYLE FOR ILLUSTRATIONS

1. Originals for illustrations should be sharp, noise-free, and of good contrast. We regret that we cannot provide drafting or art service. All original illustrations should be placed on separate pages following the text, not within the manuscript text.

2. Line drawings should be in laser printer output or in India ink on paper, or board. Use 8 1/2 by 11-inch size sheets if possible, to simplify handling of the manuscript.
3. Each figure should be mentioned in the text and numbered consecutively using Arabic numerals. Specify the desired location of each figure in the text, but place the figure itself on a separate page following the text.
4. Number each table consecutively using Arabic numerals. Please label any material that can be typeset as a table, reserving the term "figure" for material that has been drawn. Specify the desired location of each table in the text, but place the table itself on a separate page following the text. Type a brief title above each table.
5. All lettering should be large enough to permit legible reduction.
6. Photographs should be glossy prints, of good contrast and gradation, and any reasonable size.
7. Number each original on the back or at the bottom of the front.
8. Provide a separate sheet listing the captions for all figures.

PROOFING

Page proofs for articles to be included in a journal issue will be sent to the contact author for proofing, unless otherwise informed. The proofread copy should be received back by the Publisher within 72 hours.

COPYRIGHT

Upon acceptance of an article, authors will be required to sign a copyright form transferring the copyright from the authors or their employers to the Publisher.

REPRINTS

Each group of authors will be entitled to 50 free reprints of their paper. Offprints may be ordered on a form that accompanies the proofs. There are no page charges.

Nanocrystalline Metals and Oxides

Selected Properties and Applications

edited by

Philippe Knauth

MADIREL, UMR Université de Provence-SNRS, Marseille, France

Joop Schoonman

Laboratory for Inorganic Chemistry, Delft University of Technology, The Netherlands

ELECTRONIC MATERIALS: SCIENCE & TECHNOLOGY 7

Nanostructured materials have at least one dimension in the nanometer range. They became a very active research area in solid state physics and chemistry in recent years with anticipated applications in various domains, including solar cells, electronics, batteries and sensors.

Nanocrystalline metals and oxides are dense polycrystalline solids with a mean grain size below 100 nm. This book is intended to give an overview on selected properties and applications of nanocrystalline metals and oxides by leading experts in the field.

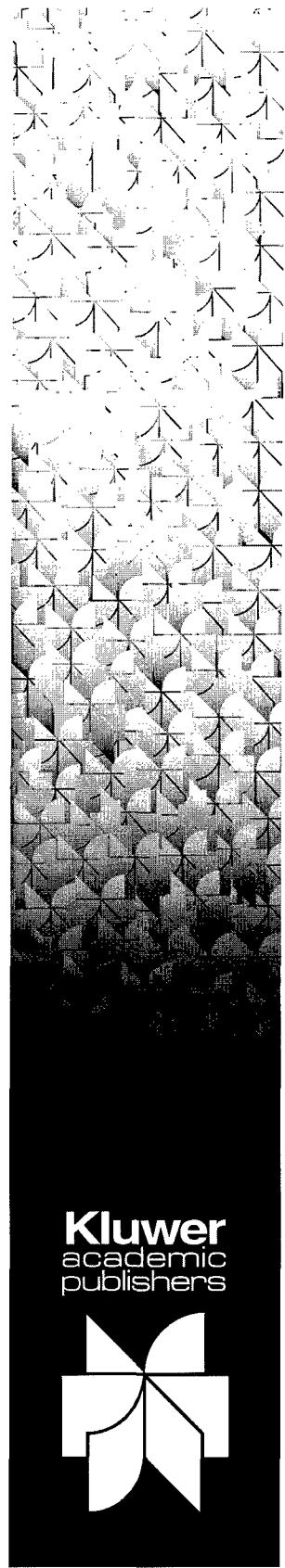
The first three chapters provide a very complete theoretical treatment of thermodynamics and atom/ion transport for nanocrystalline materials. The following chapters are experts' views on the development of experimental characterization techniques for nanocrystalline solids with emphasis on electroceramic materials.

Nanocrystalline Metals and Oxides is intended for a broad range of readers, foremost chemists, physicists and materials scientist. Theoretical physicists and chemists will certainly also profit from this book. The electroceramics and solid state ionics community are particularly addressed, given the main interests of the editors.

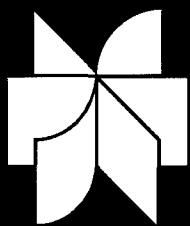
2002, 256 pp.

Hardbound, ISBN 0-7923-7627-7

Price: EUR 150.00 / GBP 93.00 / USD 135.00



Kluwer
academic
publishers



P.O. Box 322, 3300 AH Dordrecht, The Netherlands, E-mail orderdept@wkap.nl
P.O. Box 358, Accord Station, Hingham, MA 02043-0358, U.S.A., E-mail kluwer@wkap.com

Kluwer Alert

The cutting edge of research

Kluwer Alert is a free electronic notification service offered by Kluwer. It is the fastest and easiest way to receive advance notification about Kluwer's journal tables of contents, Kluwer's forthcoming books, journal issues and electronic products. Details are sent directly to your desktop, weeks ahead of publication.

Create your own unique user profile by selecting particular subject areas and specific journals listed on www.kluweralert.com. Your privacy is guaranteed!

Kluwer's email alerting system will keep you on the cutting edge of research with announcements regarding:

- ...❖ ***Journal Tables of Contents - available up to 8 weeks prior to print availability***
- ...❖ ***New books - by the top scholars in their fields***
- ...❖ ***New journals - with articles written by the world's leading scientists***
- ...❖ ***New electronic products - allowing you unprecedented flexibility and convenience in your research***
- ...❖ ***Notification of special offers at upcoming conferences***
- ...❖ ***And much more...***

With Kluwer Alert linking you directly to new products and new articles, keeping up-to-date on new information in your field of interest has never been easier.

Visit www.kluweralert.com to sign up today!



ORDER CARD—Journal of Computational Electronics
(ISSN 1569-8025)

JCEL

Institutional Rate Vol. 1 (2002)

Print or Electronic Version US \$350.00 (includes postage and handling)

Individual subscription prices are available upon request from the publisher.

Check enclosed in the amount of _____. Purchase order attached #_____

Charge my credit card account VISA MasterCard AMEX
Card no. _____ Exp. Date _____

Signature _____

Please send a free sample copy.

Name _____

Address _____

City _____ State _____ Zip _____

ORDER CARD—Journal of Computational Electronics
(ISSN 1569-8025)

JCEL

Institutional Rate Vol. 1 (2002)

Print or Electronic Version EURO 350.00 (includes postage and handling)

Individual subscription prices are available upon request from the publisher.

Check enclosed in the amount of _____. Purchase order attached #_____

Charge my credit card account VISA MasterCard AMEX
 Eurocard Diners Club
Card no. _____ Exp. Date _____

Signature _____

Please send a free sample copy.

Name _____

Address _____

City _____ Country _____ Post Code _____



NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS PERMIT NO. 148 HINGHAM, MA

POSTAGE WILL BE PAID BY ADDRESSEE



KLUWER ACADEMIC PUBLISHERS
P.O. BOX 358
ACCORD STATION
HINGHAM, MA 02018-0358

POSTAGE
STAMP
REQUIRED

KLUWER ACADEMIC PUBLISHERS
Distribution Center
P.O. BOX 322
3300 AH DORDRECHT
THE NETHERLANDS

ORDERING INFORMATION

- Prepayment is required on all orders.
- By signing this order, individual (private) subscribers declare that the subscription is for personal use only.
- All prices include postage and handling.

LIBRARY RECOMMENDATION CARD

Use this card to recommend
that your library subscribe to

Journal of Computational Electronics

ROUTE VIA INTERDEPARTMENTAL MAIL

TO: The Serials Librarian at _____

FR: _____ TITLE: _____

DEPT: _____ PHONE: _____

I recommend that our library carry a subscription for *Journal of Computational Electronics* (ISSN 1569-8025), published by Kluwer Academic Publishers.

The Institutional subscription price (including postage and handling) for 2002,
volume 1 (4 of issues) is:

Print or electronic version: US \$350.00/EURO 350.00

Signature _____ Date _____

ORDER FROM:



Kluwer Academic * Order Dept. * P.O. Box 358 * Accord Station *
Hingham, MA 02043-0358
or from

Kluwer Academic * Order Dept. * P.O. Box 322 *
3300 AH Dordrecht * The Netherlands

JOURNAL OF COMPUTATIONAL ELECTRONICS

AIMS AND SCOPE

The purpose of the *Journal of Computational Electronics* is to bring together research on all aspects of modeling and simulation of modern electronics, including optical, electronic, mechanical, and quantum mechanical aspects, as well as research on the underlying mathematical algorithms and computational details. The newer related areas of molecular and biological systems, in which the thrust is on the transport, mechanical, and optical properties, will also be covered. The Editors will emphasize advances and challenges arising from applications in multiscale problems focusing on those whose basis arises from physical and chemical sciences. A short list of the specific topics that lie within the scope of this new journal is:

- Semiconductor Devices:
 - Transport physics of ultra small structures;
 - Role of quantum effects—the transition from classical to quantum environment;
 - 2D and 3D device simulations.
- Optical Devices:
 - Semiconductor laser diodes (VCSELs and VCSEL arrays, etc.);
 - Detectors (limits of high speeds and low signal intensities, etc.);
 - Coupling to the electromagnetics.
- Process Simulation:
 - Ab initio models;
 - Many-body interactions;
 - Molecular dynamics and Monte Carlo.
- Nano-electro-mechanical systems:
 - Quantum mechanical and quantum electrodynamical forces;
 - Transition to classical approaches;
 - Coupled systems;
 - Nano-fluidics.
- Mathematical Approaches:
 - Algorithms;
 - Implementation on parallel systems;
 - Special problems with quantum mechanics.
- Correlated Areas:
 - Biological systems, ion channels, etc.;
 - Optical, mechanical and electronic interactions in molecular systems.
- Open Quantum Systems:
 - Quantum electronics;
 - Special computational problems in quantum systems;
 - Quantum computing.

JOURNAL OF COMPUTATIONAL ELECTRONICS

Volume 1, Numbers 1-2, July 2002

Contents

Proceedings of the 8th INTERNATIONAL WORKSHOP ON COMPUTATIONAL ELECTRONICS (IWCE-8),
Beckman Institute, University of Illinois, October 15-18, 2001

Editorial	<i>Umberto Ravaioli</i>	7
Eigenstate Selection in Open Quantum Dot Systems: On the True Nature of Level Broadening	<i>R. Akis, D.K. Ferry and J.P. Bird</i>	9
On the Completeness of Quantum Hydrodynamics: Vortex Formation and the Need for Both Vector and Scalar Quantum Potentials in Device Simulation	<i>John R. Barker</i>	17
On the Current and Density Representation of Many-Body Quantum Transport Theory	<i>John R. Barker</i>	23
A Space Dependent Wigner Equation Including Phonon Interaction	<i>M. Nedjalkov, H. Kosina, R. Kosik and S. Selberherr</i>	27
RTD Relaxation Oscillations, the Time Dependent Wigner Equation and Phase Noise	<i>H.L. Grubin and R.C. Buggeln</i>	33
Modeling of Shallow Quantum Point Contacts Defined on AlGaAs/GaAs Heterostructures: The Effect of Surface States	<i>G. Fiori, G. Iannaccone and M. Macucci</i>	39
Study of Noise Properties in Nanoscale Electronic Devices Using Quantum Trajectories	<i>Xavier Oriols, Ferran Martin and Jordi Suñé</i>	43

Table of Contents continued on page 3

Electronic journals at

KluwerOnline
WWW.KLUWERONLINE.NL

Contact your librarian for more information



1569-8025(200207)1:1/2;1-D